## Professor's Contact Information

|  |  |
|---|---|
| **Office Phone** | 940) 565-4769 |
| **Office Location** | BLB 312E |
| **Email Address** | Kashif.saeed@unt.edu |
| **Office Hours** | Wednesdays & Thursdays 5pm-6pm |
| **TA Information** | TA information & 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 laptop.** |
| **Course Description** | The course covers the following topics:<br>1. Hadoop Architecture and Ecosystem<br>2. HDFS and basic LINUX commands<br>3. Hive, Sqoop, Other Ecosystem tools as needed<br>4. Spark using Scala<br>5. Cloud fundamentals and cloud deployment of Big Data |
| **Student Learning Objectives** | **SLO1:** Students will be able to describe architecture and methods for storage and provision of structured and unstructured data.<br><br>**SLO2:** Students will develop competency in using Big Data platform like Hadoop, EMR, or Spark to store and retrieve big data.<br><br>**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.<br><br>**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<br>• Take notes and work on hands-on activities<br>• Good understanding of the assignments – there can be questions in the exams from the assignments<br>• 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<br>2. O'Reilly Programming Hive by Rutherglen, Wampler, and Capriolo<br>3. O'Reilly Programming Pig by Alan Gates<br>4. O'Reilly Learning Spark by Karau and Zahaia<br>5. Apache Spark and Scala documentation |
| **Software Used** | 1. Cloudera VM – will be made available by the instructor<br>2. VMware<br>https://my.vmware.com/web/vmware/free#desktop_end_user_computing/vmware_player/7_0<br>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

| Lecture | Date | Topics to be covered | Assignments |
|---|---|---|---|
| 1 | Jan 20 | • Syllabus Overview and Expectations<br>• Data Architecture concepts<br>• Infrastructure concepts<br>• Introduction to Hadoop | Install lockdown browser.<br>Install Hadoop. |
| 2 | Jan 27 | • Introduction to Hadoop<br>• Cloudera Setup overview<br>• HDFS Commands<br>• Basic LINUX commands<br>• Hands-on activities | Follow Canvas assignment section for dates for the assignments and activities |
| 3 | Feb 3 | • Sqoop Overview<br>• Import and Export commands with Sqoop<br>• Different tools and parameters in Sqoop<br>• Sqoop Hands-on<br>• Remaining topics in Sqoop | |
| 4 | Feb 10 | • Hive Overview<br>• Basic Hive and MySQL commands | |
| 5 | Feb 17 | • Hive Commands and Hands-on<br>• Importing from Sqoop into Hive<br>• Partitioning and Indexing in Hive | |
| 6 | Feb 24 | • Handling JSON data with Hive<br>• Additional advanced topics with Hive<br>• Hive Hands-on | |
| 7 | Mar 3 | • Introduction to Pig<br>• Pig Hands-on<br>• Streaming tools in Hadoop Ecosystem<br>• Overview of Flume, Kafka, Flink, etc.<br>• Introduction to Solr | |
| 8 | Mar 10 | **Midterm Exam (Testing center)**<br>**Syllabus:** Everything covered up till this point | |
| | Mar 17 | **Spring break – no class** | |
| 9 | Mar 24 | • Introduction to Scala<br>• Scala syntax | |

| | | | |
|---|---|---|---|
| | | • Control structures in Scala<br>• Functional programming in Scala<br>• Scala hands-on | |
| 10 | Mar 31 | • Introduction to Spark<br>• RDDs in Spark<br>• Spark internal concepts<br>• Spark hands-on | |
| 11 | Apr 7 | • Key-value pair RDDs<br>• Data manipulations in Spark<br>• Spark data frames | |
| 12 | Apr 14 | • Spark SQL and Spark SQL<br>• Overview of additional topics in Spark | |
| 13 | Apr 21 | • Introduction to Cloud Computing<br>• AWS Cloud Fundamentals<br>• AWS Big data and data lake | |
| 14 | Apr 28 | • GCP Cloud Fundamentals<br>• GCP Big Data and ML overview<br>• GCP Big Data<br>• Any remaining topics<br>• Exam review and discussion | |
| | May 5 | **Non-comprehensive Final Exam**<br>**Syllabus:** Everything post midterm | |

**Course Policies**

| | |
|---|---|
| **Grading** | **Grading Criteria**<br><br>A: 360+ points OR top 30% based on total enrolled (after last date to drop)<br>B: 300-359 points OR Next 40% based on total enrolled (after last date to drop)<br>C: 240-200 points OR Next 20% based on total enrolled (after last date to drop)<br>D and F: Less than 50% in exams and quizzes OR bottom 10% based on total enrolled (after last date to drop)<br><br>**Grading Points Distribution (400 total)**<br><br>Assignments – 100 points (2 @ 50 points each)<br>Activities – 50 points<br>Pop Quizzes – 50 points<br>Midterm – 100 points<br>Final – 100 points |
| **COVID-19** | Face Coverings in the class are preferred. One-on-one discussions with the professor require maintaining social distance and a face covering. This is subject to change as per the UNT and CDC guidelines. Any changes will be communicated via the instructor.<br><br>Absences: |

| | |
|---|---|
| | In-person attendance in strongly encouraged in this class.<br>While attendance is expected as outlined above, it is important for all of us to be mindful of the health and safety of everyone in our community, especially given concerns about COVID-19. Please contact me if you are unable to attend class because you are ill, or unable to attend class due to COVID-19 including symptoms, potential exposure, pending or positive test results, or if you have been given specific instructions to isolate or quarantine from a health care provider or a local authority. It is important that you communicate with me prior to being absent so I may make a decision about accommodating your request to be excused from class.<br>If you are experiencing any symptoms of COVID-19 please seek medical attention from the Student Health and Wellness Center (940-565-2333 or askSHWC@unt.edu) or your health care provider PRIOR to coming to campus. UNT also requires you to contact the UNT COVID Hotline at 844-366-5892 or COVID@unt.edu for guidance on actions to take due to symptoms, pending or positive test results, or potential exposure. While attendance is an important part of succeeding in this class, your own health, and those of others in the community, is more important. |
| **Random knowledge check** | The instructor reverses the right to check your understanding about the assignments and/or activities submitted by you. Failure to satisfactory justification on how the assignment and/or activities were completed may result in a reduced (or even a zero) grade in the assignment. |
| **Academic Integrity** | According to the UNT's academic integrity policy 06.003, "Academic Misconduct" means the intentional or unintentional action by a student to engage in behavior in the academic setting including, but not limited to: cheating, fabrication, facilitating academic misconduct, forgery, plagiarism, and sabotage.<br>Any act of academic misconduct will not be tolerated and will result in an 'F' in the class. |
| **Exams and Quizzes** | The quizzes will be conducted in the classroom. The exams will be conducted in the testing center. |
| **Make-up Exams** | There will be no make-up exams. |
| **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's 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. |
| **Exam Reviews** | Exam reviews do not 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 answer your questions, however, it is your responsibility to come prepared to the exam review class to ask questions.<br>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 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].<br><br>**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<br><br>**Emergency Evacuation Procedures for Business Leadership Building:**<br>**Severe Weather:** In the event of severe weather, all building occupants should immediately seek shelter in the designated shelter-in-place area in the building.  If unable to safely move to the designated shelter-in-place area, seek shelter in a windowless interior room or hallway on the lowest floor of the building.    All building occupants should take shelter in rooms 055, 077, 090, and the restrooms on the basement level.  In rooms 170, 155, and the restrooms on the first floor.<br>**Bomb Threat/Fire:** In the event of a bomb threat or fire in the building, all building occupants should immediately evacuate the building using the nearest exit.  Once outside, proceed to the designated assembly area.  If unable to safely move to the designated assembly area, contact on or more members of your department or unit to let them know you are safe and inform them of your whereabouts.  Persons with mobility impairments who are unable to safely exit the building should move to a designated area of refuge and await assistance from emergency responders. All building occupants should immediately evacuate the building and proceed to the south side of Crumley Hall in the grassy area, wests of parking lot 24.<br><br>**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.*

## ETHICAL ACADEMIC BEHAVIOR IN ITDS CLASSES

The UNT College of Business and the ITDS Department expect their students to behave at all times in an ethical manner.  There are at least two reasons for this.  First, ethical behavior affirms the personal value and worth of the individual.  Second, professionals in all fields (but particularly in information systems, accounting, and HR) frequently handle confidential information on behalf of their employers and clients.  Thus employers of UNT College of Business graduates expect ethical conduct from their employees because that behavior is crucial to the success of the organization.  Academic dishonesty is a major violation of ethical behavior.

Students are expected to read (https://policy.unt.edu/policy/06-003) UNT's Student Standards of Academic Integrity which defines academic dishonesty and sets out the consequences for unethical academic behavior.  Cheating and plagiarism are the most common types of academic dishonesty.

The UNT's Student Standards of Academic Integrity policy defines cheating as: *The use of unauthorized assistance in an academic exercise, including but not limited to:*
   1. *Use of any unauthorized assistance to take exams, tests, quizzes or other assessments;*
   2. *Dependence upon the aid of sources beyond those authorized by the instructor in writing papers, preparing reports, solving problems or carrying out other assignments;*
   3. *Acquisition, without permission, of tests, notes or other academic materials belonging to a faculty or staff member of the University;*
   4. *Dual submission of a paper or project, or re-submission of a paper or project to a different class without express permission from the instructor;*
   5. *Any other act designed to give a student an unfair advantage on an academic assignment.*

The university's policy defines plagiarism as the "*Use of another's thoughts or words without proper attribution in any academic exercise, regardless of the student's intent, including but not limited to:*
   1. *The knowing or negligent use by paraphrase or direct quotation of the published or unpublished work of another person without full and clear acknowledgement or citation.*
   2. *The knowing or negligent unacknowledged use of materials prepared by another person or by an agency engaged in selling term papers or other academic materials.*

Examples of academic dishonesty in an ITDS class include: copying answers from another person's paper; using notes during an exam; copying computer code from another person's work; having someone else complete your assignments or take tests on your behalf; stealing code printouts, software, or exams; recycling assignments submitted by others in prior or current semesters as your own; and copying the words or ideas of others from books, articles, reports, presentations, etc. for use as your own thoughts without proper attribution (i.e., plagiarism). It does not matter whether you received permission from the owner of the copied work; claiming the material as your own is still academic dishonesty.

The ITDS Department believes it is very important to protect honest students from unfair competition with anyone trying to gain an advantage through academic dishonesty. Academic dishonesty is not tolerated in ITDS classes, and those who engage in such behavior are subject to sanctions as outlined in the UNT's policy and/or the course syllabus. You are strongly encouraged to read the policy carefully so that you are aware of what constitutes academic dishonesty and the consequences of this unethical behavior.

By signing below, I acknowledge my responsibility to read the UNT academic dishonesty policy and the Student Standards of Academic Integrity (https://policy.unt.edu/policy/06-003); and attest that I have read and understand the statements in this document and agree to behave ethically in this class.

_____  _____
Student Name (Print)                                        Student ID No.
_____  _____
Student Signature                                          Date

---

## UNIVERSITY OF NORTH TEXAS
## Authorization to Release Assignments to Plagiarism Detection Service

Written assignments in this course will be provided to an internet-based plagiarism detection service that is not affiliated with the University of North Texas. If you sign the form, your assignments may be submitted to the service with your name or student identification number (but only if you put this number on your assignment which you should never do in this course). **If you do not sign this form, you must sign the bottom section acknowledging that it is your responsibility to make sure your name and other identifying information only appear on the coversheet of your assignments – Your cover sheet will not be included when the rest of your assignment is submitted to the service**. This authorization is only to allow the instructor to manage more efficiently the course and will expire upon the issuance of a final grade. Please sign and date the authorization form. Return the form to the instructor upon completion. You are not required to sign this authorization and you will not be penalized if you do not sign the form; however, if you do not wish to sign it you are required to complete the section at the bottom of the form.

I, _____ **[Print Name of Student],**

hereby voluntarily authorize the instructor in BCIS 4690 to disclose assignments that contain my name and/or student identification number (if for some reason I chose to include on my assignments) to an internet-based plagiarism detection service. (There is no reason you would ever put your student identification number in any assignment for this course.)

**This authorization will remain in effect from the date it assigned until a grade is assigned in this course and does not apply to any other course in which I am enrolled at the University of North Texas.**

_____
Student Signature                          Date
_____
\_\_\_\_\_

### SIGN & DATE ABOVE
### OR
### SIGN & DATE THE SECTION BELOW

_____
\_\_\_\_\_

I, _____ **[Print Name of Student],**

understand that by not agreeing to the above part of this form it is my responsibility to make sure that my name and other identifying information only appear on the coversheet of all my assignments (including team

assignments) since my instructor will submit the rest of my assignments (excluding the coversheet) to an internet-based plagiarism detection service.

_____    _____

Student Signature                                        Date

# UNT College of Business Student Ethics Statement

As a student of the UNT College of Business, I will abide by all applicable policies of the University of North Texas, including the Student Standards of Academic Integrity, the Code of Student Conduct and Discipline and the Computer Use Policy.  I understand that I am responsible reviewing the policies as provided by link below before participating in this course.  I understand that I may be sanctioned for violations of any of these policies in accordance with procedures as defined in each policy.

I will not engage in any acts of academic dishonesty as defined in the Student Standards of Academic Integrity, including but not limited to using another's thoughts or words without proper attribution (plagiarism) or using works in violation of copyright laws. I agree that all assignments I submit to the instructor and all tests I take shall be performed solely by me, except where my instructor requires participation in a group project in which case I will abide by the specific directives of the instructor regarding group participation.

While engaged in on-line coursework, I will respect the privacy of other students taking online courses and the integrity of the computer systems and other users' data.  I will comply with the copyright protection of licensed computer software.  I will not intentionally obstruct, disrupt, or interfere with the teaching and learning that occurs on the website dedicated to this course through computer "hacking" or in any other manner.

I will not use the university information technology system in any manner that violates the UNT nondiscrimination and anti-sexual harassment policies.  Further, I will not use the university information technology system to engage in verbal abuse, make threats, intimidate, harass, coerce, stalk or in any other manner which threatens or endangers the health, safety or welfare of any person.  Speech protected by the First Amendment of the U.S. Constitution is not a violation of this provision, though fighting words and statements that reasonably threaten or endanger the health and safety of any person are not protected speech.

Student Standards of Academic Integrity
http://policy.unt.edu/sites/default/files/untpolicy/pdf/7-Student_Affairs-Academic_Integrity.pdf

Code of Student Conduct and Discipline
http://conduct.unt.edu/sites/default/files/pdf/code_of_student_conduct.pdf

Computer Use Policy: http://policy.unt.edu/policy/3-10

By signing below, I acknowledge my responsibility to read the UNT academic dishonesty policy and the Student Standards of Academic Integrity (https://policy.unt.edu/policy/06-003); and attest that I have read and understand the statements in this document and agree to behave ethically in this class.

_____    _____

Student Name (Print)                                               Student ID No.

_____    _____

Student Signature                                            Date