CSE6331: Cloud Computing (Spring 2021)

Instructor: Leonidas Fegaras
Office Number: ERB 653 (Engineering Research Bldg)
Email Address: fegaras@cse.uta.edu
Faculty Profile: https://www.uta.edu/profiles/leonidas-fegaras
Office Hours: Tuesday and Thursday 5:30-7:00pm
Section Information: CSE 6331-005
Time and Place of Class Meetings: TuTh 7:00-8:20pm

Mode of Teaching:

Online synchronous: Lectures and exams will be online. There is no requirement for students to come to campus. The exam will be online and is required at a specific day and time. All lectures will be held live through Microsoft Teams at a scheduled day and time (TuTh 7:00-8:20pm). Public chat will be available during these live sessions. The video recording of each lecture will be uploaded on Canvas after the class. Students are strongly encouraged but not required to attend the live sessions, but if they miss a live session, they are required to watch the video of the session. Office hours will be held online on Microsoft Teams. Projects will be submitted on Canvas. You will need to use the Respondus Lockdown Browser to take the exam. You can take the exam on a Windows 10 or a Mac with a webcam. The webcam is required and must be on during the exam. The Lockdown Browser does not work on Linux. It is your responsibility to find a PC or a MAC with a webcam to take the exam.

Description of Course Content:
The focus of this course is on data management techniques and tools for storing and analyzing very large amounts of data. Topics that will be covered include: cloud computing; virtualization; distributed file systems; large data processing using Map-Reduce; data modeling, storage, indexing, and query processing for big data; key-value storage systems, columnar databases, NoSQL systems (Cassandra, BigTable, HBase, MongoDB); big data technologies and tools (Hive, Pig, Spark, Flink); large-scale stream processing systems (Storm, Spark Streaming); data analytics frameworks (Mahout); big data applications, including graph processing, recommendation systems, machine learning, clustering, classification, prediction, and stream mining.

Student Learning Outcomes: Upon successful completion of this course, students:

- will gain an in-depth understanding of many theoretical and practical aspects of Big Data management and analysis;
- will be able to make use of current technologies to design highly scalable systems that can process and analyze Big Data for a variety of scientific, social, and environmental challenges;
- will acquire knowledge and skills in designing new generation of scalable algorithms and using new data management technologies to extract, interpret, and learn from very large data sets.
Prerequisites:
Prerequisites: CSE 3330/CSE 5330 (Database Systems I) or equivalent. Students are expected to have a working knowledge of Java and some knowledge of SQL. Students without adequate preparation are at substantial risk of failing this course.

Required Textbooks and Other Course Materials:
There is no required textbook for this course but students are expected to read many online tutorials and references (links will be given out in class).

Descriptions of major assignments and examinations:
There will be eight programming assignments and a final exam.

Grading:
The final grade will be based on
- 60% 8 small programming assignments (7.5% each)
- 40% final exam
Final grades will be assigned according to the following scale:

A: score \(\geq 90\), B: \(80 \leq \text{score} < 90\), C: \(70 \leq \text{score} < 80\), D: \(60 \leq \text{score} < 70\), F: \(\text{score} < 60\).

Sometimes, lower cutoff points are used for the final grades, depending on the overall performance of the class. Students are expected to keep track of their performance throughout the semester and seek guidance from available sources (including the instructor) if their performance drops below satisfactory levels.

Exam:
The final exam is open notes and books. The final exam will be online and is required at a specific day and time. You will need to use the Respondus Lockdown Browser to take the exam. You can take the exams on a Windows 10 or a Mac with a webcam. The webcam is required and must be on during the exam. The Lockdown Browser does not work on Linux. It is your responsibility to find a PC or a MAC with a webcam to take the exam. No makeup exam will be given unless there is a justifiable reason (such as illness, sickness or death in the family). If you miss the exam and you can prove that your reason is justifiable, you should arrange with the instructor to take the makeup exam within a week from the regular exam time. For any other case, you will get a zero grade for the missed exam.

Programming Assignments:
There will be 8 programming assignments. The programming assignments must be done individually. Details will be given out in class. Late assignments will be marked 20 points off per day (out of 100 max). So, there is no point submitting a project more than 4 days late! This penalty cannot be waived, unless there was a case of illness or other substantial impediment beyond your control, with proof in documents from the school.

Technology Requirements:
You will need a Windows 10 PC/laptop or a Mac to take the final exam. Your PC/laptop must have a webcam, which must be on during the exam. The webcam is required and must be on during the exams. The Lockdown Browser does not work on Linux. It is your responsibility to find a PC or a MAC with a webcam to take the exams. The projects can be done on any PC/laptop, such as on Linux, Windows, or Mac.

Software:
Most programming will be done in Java but some will be done in Scala, Pig, and Hive.
Students are expected to have a working knowledge of Java and some basic knowledge of SQL. Students will develop their programs on the cloud computing platform **SDSC Comet**, which is part of **XSEDE**, supported by NSF. Students will get a free account but there will be usage limits. Detailed instructions will be given out in class. The programming assignments will be related to HBase, Hadoop Map-Reduce, Spark, Flink, Hive, Pig, Storm, etc.

**How to do Well in this Course:**
Students who get the most out of this course will be the ones who put in the most effort. If you want to do well, attend all the online lecture sessions, read the assigned reading material, and start early on your programming assignments. If you are having difficulty, the instructor and the GTA will be more than happy to help you. In addition to regular office hours, the best way of communication with the instructor or the GTA is through email. If you can not make it to the scheduled office hours but really need help, contact one of us for an appointment.

**Institution Information:**
UTA students are encouraged to review the below institutional policies and informational sections and reach out to the specific office with any questions. To view this institutional information, please visit the [Institutional Information page](#), which includes the following policies among others:
- Drop Policy
- Disability Accommodations
- Title IX Policy
- Academic Integrity
- Student Feedback Survey
- Final Exam Schedule

**Mandatory Face Covering Policy:**
All students and instructional staff are required to wear facial coverings while they are on campus, inside buildings and classrooms. Students that fail to comply with the facial covering requirement will be asked to leave the class session. If students need masks, they may obtain them at the Central Library, the E.H. Hereford University Center’s front desk or in their department. Students who refuse to wear a facial covering in class will be asked to leave the session by the instructor; and, if the student refuses to leave, they may be reported to UTA's Office of Student Conduct.

**Attendance:**
The University of Texas at Arlington, taking attendance is not required but attendance is a critical indicator of student success. Each faculty member is free to develop his or her own methods of evaluating students’ academic performance, which includes establishing course-specific policies on attendance. As the instructor of this section, I allow students to attend class sessions at their own discretion. However, while UT Arlington does not require instructors to take attendance in their courses, the U.S. Department of Education requires that the University have a mechanism in place to mark when Federal Student Aid recipients “begin attendance in a course.” UT Arlington instructors will report when students begin attendance in a course as part of the final grading process. Specifically, when assigning a student a grade of F, faculty report must the last date a student attended their class based on evidence such as a test, participation in a class project or presentation, or an engagement online via Canvas. This date is reported to the Department of Education for federal financial aid recipients.
Tentative Course Schedule:

- Cloud computing fundamentals:
  - high-performance computing
  - parallel, distributed, cluster, grid computing
  - cloud computing architecture and management
  - cloud service models: IaaS, PaaS, SaaS, STaaS
  - cloud technology: virtualization, hypervisors

- Introduction to Big-Data technologies and tools

- Big-Data storage and indexing:
  - distributed file systems: HDFS
  - key-value storage systems
  - columnar databases
  - NoSQL systems: Cassandra, BigTable, HBase, MogoDB

- Programming models:
  - Hadoop Map-Reduce
  - Apache Spark
  - Apache Flink

- Big-Data languages: Hive, Pig, Spark SQL/DataFrame, Cascading

- Languages for graph processing: Pregel, Giraph, GraphX

- Data analytics frameworks: Mahout, MLlib

- Large-scale stream processing systems: Storm, Spark Streaming

- Big Data data analysis applications:
  - PageRank
  - clustering
  - classification
  - recommendation systems

As the instructor for this course, I reserve the right to adjust this schedule in any way that serves the educational needs of the students enrolled in this course. - Leonidas Fegaras.

Emergency Phone Numbers:
In case of an on-campus emergency, call the UT Arlington Police Department at 817-272-3003 (non-campus phone), 2-3003 (campus phone). You may also dial 911. Non-emergency number 817-272-3381

Last modified: 11/09/2020 by Leonidas Fegaras