

CSE6331: Cloud Computing (Spring 2019)

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 3:30-5:00pm
Section Information: CSE 6331-004 and 6331-005
Time and Place of Class Meetings: TuTh 2:00-3:20 (6331-004) and 5:30-6:50pm (6331-005)

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, MogoDB); 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, a midterm exam, and a final exam.

Attendance:

At The University of Texas at Arlington, taking attendance is not required but attendance is a critical indicator in 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 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 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 Blackboard. This date is reported to the Department of

Education for federal financial aid recipients.

Grading:

The final grade will be based on

- 48% 8 small programming assignments (6% each)
- 20% midterm exam
- 32% final exam (comprehensive)

Final grades will be assigned according to the following scale:

A: score ≥ 90 , B: $80 \leq \text{score} < 90$, C: $70 \leq \text{score} < 80$, D: $60 \leq \text{score} < 70$, F: 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.

Exams:

Both exams are open notes (all notes must be securely bound in one notebook). The final exam will cover the material from the first lecture up to and including the last lecture. Once the exam grades are posted, you will have 10 business days to dispute your grade and get your exam re-evaluated. No re-evaluation will be entertained after the 10 day period. No makeup exams will be given unless there is a justifiable reason (such as illness, sickness or death in the family). If you miss an 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.

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 lectures, 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't make it to the scheduled office hours but really need help, contact one of us for an appointment.

Drop Policy:

Students may drop or swap (adding and dropping a class concurrently) classes through self-service in MyMav from the beginning of the registration period through the late registration period. After the late registration period, students must see their academic advisor to drop a class or withdraw. Undeclared students must see an advisor in the University Advising Center. Drops can continue through a point two-thirds of the way through the term or session. It is the student's responsibility to officially withdraw if they do not plan to attend after registering. Students will not be automatically dropped for non-attendance. Repayment of certain types of financial aid administered through the University may be required as the result of dropping classes or withdrawing. For more information, contact the Office of Financial Aid and Scholarships

(<http://www.uta.edu/aao/fao/>).

Disability Accommodations:

UT Arlington is on record as being committed to both the spirit and letter of all federal equal opportunity legislation, including The Americans with Disabilities Act (ADA), The Americans with Disabilities Amendments Act (ADAAA), and Section 504 of the Rehabilitation Act. All instructors at UT Arlington are required by law to provide "reasonable accommodations" to students with disabilities, so as not to discriminate on the basis of disability. Students are responsible for providing the instructor with official notification in the form of a letter certified by the Office for Students with Disabilities (OSD). Only those students who have officially documented a need for an accommodation will have their request honored. Students experiencing a range of conditions (Physical, Learning, Chronic Health, Mental Health, and Sensory) that may cause diminished academic performance or other barriers to learning may seek services and/or accommodations by contacting: The Office for Students with Disabilities, (OSD) www.uta.edu/disability or calling 817-272-3364. Information regarding diagnostic criteria and policies for obtaining disability-based academic accommodations can be found at www.uta.edu/disability. Counseling and Psychological Services, (CAPS) www.uta.edu/caps/ or calling 817-272-3671 is also available to all students to help increase their understanding of personal issues, address mental and behavioral health problems and make positive changes in their lives.

Non-Discrimination Policy:

The University of Texas at Arlington does not discriminate on the basis of race, color, national origin, religion, age, gender, sexual orientation, disabilities, genetic information, and/or veteran status in its educational programs or activities it operates. For more information, visit uta.edu/eos.

Title IX Policy:

The University of Texas at Arlington ("University") is committed to maintaining a learning and working environment that is free from discrimination based on sex in accordance with Title IX of the Higher Education Amendments of 1972 (Title IX), which prohibits discrimination on the basis of sex in educational programs or activities; Title VII of the Civil Rights Act of 1964 (Title VII), which prohibits sex discrimination in employment; and the Campus Sexual Violence Elimination Act (SaVE Act). Sexual misconduct is a form of sex discrimination and will not be tolerated. For information regarding Title IX, visit www.uta.edu/titleIX or contact Ms. Jean Hood, Vice President and Title IX Coordinator at (817) 272-7091 or jmhood@uta.edu.

Academic Integrity:

Students enrolled all UT Arlington courses are expected to adhere to the UT Arlington Honor Code: I pledge, on my honor, to uphold UT Arlington's tradition of academic integrity, a tradition that values hard work and honest effort in the pursuit of academic excellence. I promise that I will submit only work that I personally create or contribute to group collaborations, and I will appropriately reference any work from other sources. I will follow the highest standards of integrity and uphold the spirit of the Honor Code. UT Arlington faculty members may employ the Honor Code as they see fit in their courses, including (but not limited to) having students acknowledge the honor code as part of an examination or requiring students to incorporate the honor code into any work submitted. Per UT System Regents' Rule 50101, Â§2.2, suspected violations of university's standards for academic integrity (including the Honor Code) will be referred to the Office of Student Conduct. Violators will be disciplined in accordance with University policy, which may result in the student's suspension or expulsion from the University. Additional information is available at <https://www.uta.edu/conduct/>.

Electronic Communication:

UT Arlington has adopted MavMail as its official means to communicate with students about important deadlines and events, as well as to transact university-related business regarding financial aid, tuition, grades, graduation, etc. All students are assigned a MavMail account and are responsible for checking the inbox regularly. There is no additional charge to students for using this account, which remains active even after graduation. Information about activating and using MavMail is available at <http://www.uta.edu/oit/cs/email/mavmail.php>.

Campus Carry:

Effective August 1, 2016, the Campus Carry law (Senate Bill 11) allows those licensed individuals to carry a concealed handgun in buildings on public university campuses, except in locations the University establishes as prohibited. Under the new law, openly carrying handguns is not allowed on college campuses. For more information, visit <http://www.uta.edu/news/info/campus-carry/>.

Student Feedback Survey:

At the end of each term, students enrolled in face-to-face and online classes categorized as "lecture," "seminar," or "laboratory" are directed to complete an online Student Feedback Survey (SFS). Instructions on how to access the SFS for this course will be sent directly to each student through MavMail approximately 10 days before the end of the term. Each student's feedback via the SFS database is aggregated with that of other students enrolled in the course. Students' anonymity will be protected to the extent that the law allows. UT Arlington's effort to solicit, gather, tabulate, and publish student feedback is required by state law and aggregate results are posted online. Data from SFS is also used for faculty and program evaluations. For more information, visit <http://www.uta.edu/sfs>.

Final Review Week:

For semester-long courses, a period of five class days prior to the first day of final examinations in the long sessions shall be designated as Final Review Week. The purpose of this week is to allow students sufficient time to prepare for final examinations. During this week, there shall be no scheduled activities such as required field trips or performances; and no instructor shall assign any themes, research problems or exercises of similar scope that have a completion date during or following this week unless specified in the class syllabus. During Final Review Week, an instructor shall not give any examinations constituting 10% or more of the final grade, except makeup tests and laboratory examinations. In addition, no instructor shall give any portion of the final examination during Final Review Week. During this week, classes are held as scheduled. In addition, instructors are not required to limit content to topics that have been previously covered; they may introduce new concepts as appropriate.

Emergency Exit Procedures:

Should we experience an emergency event that requires us to vacate the building, students should exit the room and move toward the nearest exit, which is located next to the classroom. When exiting the building during an emergency, one should never take an elevator but should use the stairwells. Faculty members and instructional staff will assist students in selecting the safest route for evacuation and will make arrangements to assist individuals with disabilities.

Web Page:

<https://lambda.uta.edu/cse6331/>

Please visit this web page often; it will contain the reading assignments, project description, class notes, etc.

Course Tentative 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: 12/17/2018 by [Leonidas Fegaras](#)