Cloud Computing for Data Analysis
ITCS 6190/8190 – Fall 2026

Welcome to **ITCS 6190/8190 – Cloud Computing for Data Analysis**! This is a challenging course that encompasses substantial technical content and programming. Consistent study, timely completion of assignments and projects, class attendance, and seeking assistance when necessary are the keys that will lead you to great success in this course.

This syllabus contains the policies and expectations established for the course. Please read the entire syllabus carefully before continuing in this course. These policies and expectations are intended to create a productive learning atmosphere for all students. Unless you are prepared to abide by these policies and expectations, you risk losing the opportunity to participate further in the course. Any modifications will be communicated through in-class announcements and/or updates on Canvas.

Course Information

Course: ITCS 6190/8190 – Cloud Computing for Data Analysis
Section numbers: ITCS 6190-U01 / ITCS 8190-U01
Cross-listed courses: ITCS 6190 and ITCS 8190 are offered as a combined section. This syllabus applies to both course numbers.
Semester and year: Fall 2026
Syllabus date: July 27, 2026

Course Description

This is a foundational course on cloud computing technology for data-intensive applications. The course provides students with essential knowledge and practical skills on scalable and efficient data analysis. The topics were thoughtfully selected to guide students from fundamental concepts to more advanced aspects, providing a solid understanding of the domain. The course is grounded in the Apache Software Foundation's ecosystem, which hosts a vast and diverse set of open-source projects across multiple domains. In particular, the course emphasizes the Spark framework for large-scale data processing, SQL analytics, streaming, and machine learning. Other components of the ecosystem, such as Hadoop, are introduced briefly for context. In addition to open-source technologies, the course includes hands-on experience with Amazon Web Services (AWS). This leading cloud service provider offers a wide array of tools for storage, computing power, machine learning, and big data processing. Together, Spark and AWS provide students with both the theoretical foundation and the practical expertise needed to design and implement scalable cloud-based data analytics solutions.

Apache Spark is an open-source distributed computing system designed for fast and large-scale data processing, offering improved performance and ease of use compared to traditional MapReduce approaches. While Hadoop, based on Google's MapReduce technology, has historically provided a foundation for distributed storage and processing, Spark has emerged as the more flexible and efficient alternative for modern data analytics. Both frameworks influenced the evolution of cloud-based solutions, and today, many providers, including Amazon Web Services (AWS), offer managed services that integrate or support Spark to simplify large-scale data processing and analytics.

This course employs a balanced approach, combining concepts with hands-on work. Students will apply the principles they have learned to design and implement data analysis jobs using cloud computing technology.

Emphasis is on collaborative learning, with opportunities for group work, discussions, and weekly check-ins with the instructor or TAs.

Credit hours: 3

Prerequisites

Familiarity with Java, Python, SQL, Linux, Data Structures, and ML; good programming skills and a solid computer science background.
Required: ITCS 6114 or permission from the department.
There are no co-requisites for this course.

Learning Outcomes

Upon successful completion of this course, students will be able to:

- Understand and explain the principles of cloud computing, including virtualization, containerization, and distributed data processing models.
- Apply practical skills to design and implement large-scale data analysis pipelines using Apache Spark and related frameworks.
- Leverage cloud infrastructure (e.g., AWS) to deploy, manage, and scale data-intensive applications securely and efficiently.
- Utilize modern data processing techniques such as Spark SQL, Structured APIs, streaming analytics, and MLlib for batch, streaming, and machine learning workloads.
- Collaborate effectively in teams to design, implement, and evaluate cloud-based data solutions, ensuring shared responsibility and balanced contributions.
- Critically assess trade-offs between different cloud services, architectures, and frameworks for real-world big data challenges.

Instructor and Meeting Information

Instructor name and contact information, teaching assistant details, and the course meeting time and location are provided to enrolled students in Canvas. Under the UNC Charlotte Academic Procedure on syllabus requirements, this information is not required in the publicly posted syllabus.

Textbooks and Course Materials

Thomas Erl, Eric Monroy, "Cloud Computing: Concepts, Technology, Security, and Architecture", 2nd Edition, ISBN-13: 978-0138052256, Pearson, 2023.
Jules Damji, Brooke Wenig, Tathagata Das, Denny Lee, "Learning Spark: Lightning-Fast Data Analytics", 2nd Edition, ISBN-13: 978-1492050049, O'Reilly Media, 2020.
(optional) Jure Leskovec, Anand Rajaraman, Jeffrey David Ullman, "Mining of Massive Datasets", 3rd Edition, ISBN: 978-1108476348, Cambridge University Press, 2020.

Required for purchase: the Erl & Monroy and the Damji et al. texts listed above. The Leskovec et al. text is optional and is not required for purchase.

All other course materials are provided at no cost. Lecture slides, readings, and assignment descriptions are distributed through Canvas. The software used in this course (Apache Spark, Python, Git/GitHub, and Canvas) is free of charge. No lab supplies, equipment, course packs, or subscription services are required for purchase.

Cloud computing accounts: all AWS work required for this course can be completed within the AWS Free Tier, at no cost to students. Students create their own AWS account and are responsible for shutting down resources when they finish an activity so that their usage stays within Free Tier limits.

Course Topics

Getting Started: Cloud concepts and models, virtualization, containerization, and introduction to Hadoop (HDFS and MapReduce).
Spark: Big data analytics and Spark, Spark Structured APIs, Spark SQL and DataFrames, streaming analytics with Spark, and Machine Learning with MLlib.
AWS: AWS concepts and AWS services.

Diverse Scholarly Perspectives

This course engages diverse scholarly perspectives to develop critical thinking, analysis, and debate. The assignment of a reading, dataset, tool, vendor, or other course material does not imply the instructor's or the University's endorsement of the views, products, or services it contains. Students are expected to examine and evaluate all course materials critically.

Format of (most) Classes

Each class follows a structured format designed to maximize preparation, engagement, and practice.

- **Before class (mandatory):** Students are expected to review the recommended readings, check the lecture slides, and watch any assigned videos. This preparation ensures that students arrive ready to engage actively with the material.
- **Class (starting sharply at the scheduled start time):** Each session begins with a short quiz to check preparation and understanding of pre-class materials. The instructor then introduces the topic, followed by a class discussion, often grounded in a case study. After a short break, the second part of class is devoted either to a hands-on activity or a project check-in with the instructor/TAs.
- **After class:** Students are expected to submit the day's hands-on activity (if applicable) or update project action items. These post-class tasks reinforce learning and ensure steady progress on team projects.

The schedule may be adjusted slightly during the semester. Students should check Canvas regularly for updates and announcements.

Grading

Students will be evaluated through a combination of quizzes, hands-on activities, assignments, a course project, and exams.

10-minute In-class Quizzes (6 points) - At the beginning of most classes, students complete a short 10-minute quiz, which collectively accounts for 6 points out of 100. These quizzes are designed to assess students' understanding of the preparation materials provided in advance. The goal is to encourage consistent study habits and readiness for in-class activities. Makeup quizzes will not be permitted unless prior arrangements have been made or exceptional circumstances are documented and approved.

In-class Hands-on (14 points) - Throughout the semester, students will participate in hands-on activities directly related to the topics discussed in class. These activities, which account for 14 points out of the total 100, will be conducted as group exercises, with ad hoc groups of 2 or 3 students. The outcome of each activity must be submitted on Canvas before the end of the day to receive credit. Points will be awarded based on the quality of the submission, as well as attendance and punctuality, with attendance being collected at the beginning of each class. This ensures that students who arrive on time and actively participate in the hands-on activities are recognized for their efforts.

Assignments (10 points) - Students will complete five individual assignments, which together account for 10 points out of the total 100. Each assignment will be made available in advance, giving students sufficient time to prepare and complete the work. Assignments are designed to reinforce class concepts and require students to apply their knowledge independently. All submissions must be uploaded to Canvas before the specified deadline to receive credit. Timely submission is essential, as late work will not be accepted without prior approval under exceptional circumstances.

Course Project (40 points) - Students will complete one major group project, worth 40 points out of the total 100. The project focuses on designing and implementing an end-to-end big data analytics pipeline using Apache Spark. Each team will select a dataset, define meaningful analytical or predictive questions, and build a pipeline that integrates structured APIs for ingestion, transformations, and aggregations, Spark SQL for complex queries, streaming for real-time or simulated ingestion, and MLlib for machine learning tasks with appropriate evaluation metrics. All development and collaboration will be conducted through GitHub, in accordance with professional practices. Groups will consist of five students unless otherwise approved, and all members are expected to contribute meaningfully and to be familiar with the complete pipeline, not only with their individual contributions. Teams will check in weekly with TAs or the instructor and document their progress in GitHub Issues. Each group will also present their project in class, demonstrating the pipeline, highlighting key findings, and reflecting on lessons learned. Grading will consider multiple dimensions: the initial proposal, which defines the dataset and project goals; consistent progress tracked through milestones and check-ins; the technical quality of the implementation; the clarity and depth of analysis and insights; and the final presentation of the project.

Midterm (10 points) + Final Exam (20 points) - The course includes two major exams: a midterm and a final. The midterm, worth 10 points, will take place during class on a date announced in class and posted on Canvas. The final exam, worth 20 points, will be held during the University's final examination period, at the date and time set by the Registrar and posted on Canvas. Both exams will be closed-book and conducted on paper, requiring students to rely solely on their knowledge and understanding of the course material. These exams are designed to assess students' comprehension of key concepts and their ability to apply what they have learned throughout the semester.

Standard grading:

- 100%–90%: A
- <90%–80%: B
- <80%–70%: C
- <70%: U

ITCS 6190 and ITCS 8190 are graduate courses. Under the UNC Charlotte Academic Policy: Grading (Graduate), graduate courses are graded A, B, C, and U (unsatisfactory); the grade of D is not assigned.

In-class Hands-on, Assignments & Project Submissions

Canvas will be used for assignment submissions, while the course project will be managed and submitted via GitHub. Students should regularly check Canvas for important dates, materials, and class announcements.

Late submissions of assignments will result in a grade reduction unless explicitly authorized by the instructor. The penalty will be as follows: 20% after one day, 50% after two days, and 100% after three days. In other words, submitting more than 3 days after the deadline will result in 0 credit.

Students may request regrading of their work. Requests for assignment regrading must be submitted via Canvas, while regrading of exams or the course project must be requested by email to the instructor. Group work will be assessed based on the overall output and each member's contributions.

Policies

I. Course Materials: All lectures and course material will be available in Canvas. Lectures and course materials, including presentations, assignments, exams, outlines, and similar materials, are protected by copyright. You are encouraged to take notes and make copies of course materials for your educational use. However, you may not, nor may you knowingly allow others to reproduce or distribute lecture notes and course materials publicly without my express written consent. Note that the syllabus itself is a directed work owned by the University and is a public record under Chapter 132 of the North Carolina General Statutes.

II. Classroom Conduct: We will conduct this class in an atmosphere of mutual respect. Active participation in class discussions is greatly encouraged. The conflict of ideas is encouraged and welcome. The orderly questioning of the ideas of others, including mine, is similarly welcome.

III. Attendance and Absences: Students are expected to attend every class and remain in class for the duration of the session. Attendance is required as part of the graded component of the In-class Hands-on category (14 points). Missing a class also means forfeiting that day's in-class quiz (6 points).

IV. Use of Generative Artificial Intelligence (AI): In this course, students are permitted to use generative artificial intelligence (AI) tools to support their work. To maintain academic integrity, students must disclose any AI-generated material they use and properly attribute it, including in-text citations, quotations, and references. Students should also include an acknowledgment statement in their assignments.

V. Academic Integrity & Plagiarism: All students are required to read and abide by the Code of Student Academic Integrity. Commercial course-material and homework-help services such as CourseHero and Chegg are prohibited.