

Cloud Computing for Data Analysis

ITCS 6190 – Summer 2026

Welcome to *ITCS 6190 – 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 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.

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.

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.
- Critically assess trade-offs between different cloud services, architectures, and frameworks for real-world big data challenges.

Location and Time

This course is offered fully online and asynchronously (May 18 – June 22, 2026). All materials, quizzes, and hands-on activities are available on Canvas. Additional online sessions to help students with course content may be scheduled as needed and announced on Canvas. The course concludes with a required live presentation session on June 22–23 (exact time TBD).

Instructor

Marco Vieira

Email: marco.vieira@charlotte.edu

Office Hours: Thursday, 4:00 pm to 5:00 pm + on demand (just send an email)

Zoom: <https://charlotte-edu.zoom.us/my/mvieira1>

Teaching Assistants (*see Canvas for details*)**Textbook(s)**

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.

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.

Format of (most) Classes

This course is fully asynchronous. Each week covers two lectures, each following a consistent rhythm:

- **Prepare:** Review the recommended readings, watch the assigned lecture videos, and check the slides. Preparation is essential before attempting the quiz.
- **Quiz:** Complete the timed quiz on *Canvas* within the weekly window. Quizzes assess understanding of the preparation materials and must be completed before the hands-on activity.
- **Hands-on:** Complete the individual hands-on activity and submit on *Canvas* before the end of the week's lecture. Work on the course project incrementally throughout the week.

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, a course project, and a live presentation.

10-minute Quizzes (10 points) - For each lecture, students complete a short 10-minute timed quiz on *Canvas*, which collectively accounts for 10 points out of 100. Quizzes are designed to assess understanding of the preparation materials and must be completed within the weekly window published in the course schedule. Makeup quizzes will not be permitted unless prior arrangements have been made or exceptional circumstances are documented and approved.

Hands-on Activities (30 points) - Each lecture includes an individual hands-on activity directly related to the lecture topic. These activities, which collectively account for 20 of the 100 points, must be completed individually and submitted on *Canvas* before the end of the week's lecture. Points will be awarded based on the quality and completeness of the submission.

Course Project (30 points) - Students will complete one project, individually or in pairs, worth 30 of the 100-point total. The project focuses on designing and implementing an end-to-end big data analytics pipeline using Apache Spark. Students will select a dataset from a set of pre-approved options, 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 will be conducted through GitHub, in accordance with professional practices. Students will submit a written report as part of the final deliverable. Grading will consider the initial proposal, which defines the dataset and project goals; the technical quality of the implementation; the clarity and depth of analysis and insights; and the final presentation. Pairs are expected to contribute equally, and both members must be able to speak to the full pipeline during the live presentation.

Live Presentation + Concept Defense (30 points) - The course concludes with a required live synchronous session (June 22–23; exact time TBD), which accounts for 36 of the total 100 points. Each student presents their individual pipeline project in two parts: a pipeline demo (~10 minutes), walking through the implementation and key design decisions, and a concept defense (~7 minutes) in which the instructor asks questions drawn from a topic list published at the start of the course. These questions may cover any concept addressed in the course — including cloud computing models, virtualization, containerization, Hadoop, Spark internals, SQL and DataFrames, streaming, MLlib, and AWS services — and are not limited to what the student chose to implement in their project. Students are expected to demonstrate a solid understanding of the full course material, not only the components of their own pipeline. Two time slots will be offered; attendance is mandatory. Students who cannot attend either slot must contact the instructor in advance to arrange an accommodation.

Standard grading -

100%-90%: A

<90%-80%: B

<80%-70%: C

<70%: F

Hands-on & 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 will result in a grade reduction unless explicitly authorized by the instructor. Given the compressed pace of the summer course, the penalty will be as follows: 10% per day late, with 0 credit after 5 days. No late submissions will be accepted after the last day of the course.

Students may request regrading of their work. Requests for regrading of hands-on activities must be submitted via *Canvas*, while regrading of the project or presentation must be requested by email to the instructor.

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. This includes providing materials to commercial course material suppliers such as CourseHero, Chegg, and other similar services. Students who publicly distribute or display or help others publicly distribute or display copies or modified copies of an instructor's course materials may violate University Policy 406, The Code of Student Responsibility, or University Policy 407, Code of Student Academic Integrity. Similarly, you own copyright in your original papers and exam essays.

II. Online Conduct

All students and the instructor are expected to engage with each other respectfully. Active participation in online discussions is greatly encouraged. Each of us may have strongly differing opinions on the various topics discussed. The conflict of ideas is encouraged and welcome. The orderly questioning of others' ideas, including my own, is similarly welcome. However, I will exercise my responsibility to ensure that all interactions (whether on discussion boards, during the live session, or via email) remain constructive and collegial. You should expect that if your conduct seriously disrupts the atmosphere of mutual respect I expect in this course, you will not be permitted to participate further.

III. Attendance and Engagement

This is a fully asynchronous course, and students are expected to engage consistently with all course materials each week. Failure to complete quizzes, hands-on activities, or project milestones on time may impact your ability to achieve the course objectives and affect your grade. An absence from a deadline, excused or unexcused, does not relieve a student of any course requirement.

Attendance at the live presentation session (June 22–23) is mandatory. Students who cannot attend either of the two offered time slots must contact the instructor in advance to arrange an accommodation. Failure to participate in the live session without prior arrangement will result in a grade of zero for the Live Presentation + Concept Defense component.

The instructor has the authority to excuse a student's absence from the live session and to grant academic accommodations (e.g., an alternative time slot, an extended deadline) under exceptional circumstances. University-sanctioned events or activities are considered excused absences. For more information, see provost.charlotte.edu/policies-procedures/academic-policies-and-procedures/course-attendance-and-participation.

IV. Non-Discrimination

All students and the instructor are expected to engage with each other respectfully. Unwelcome conduct directed toward another person based upon that person's actual or perceived race; color; religion (including belief and non-belief); sex; sexual orientation; gender identity; age; national origin; physical or mental disability; veteran status; genetic information; or for any other reason, may constitute a violation of University Policy 501, Nondiscrimination. Any student suspected of engaging in such conduct will be referred to the Office of Civil Rights & Title IX.

V. University Policy on Withdrawals

Students are expected to complete all courses for which they are registered at the close of the add/drop period. If you are concerned about your ability to succeed in this course, it is important to schedule an appointment with me as soon as possible. The University policy on withdrawal allows students only a limited number of opportunities to withdraw from courses. It is important for you to understand the financial and academic consequences of course withdrawal. See: provost.charlotte.edu/policies-procedures/academic-policies-and-procedures/withdrawal-and-cancellation-enrollment-policy

VI. Use of Generative Artificial Intelligence (AI)

In this course, students may use generative artificial intelligence (AI) tools such as ChatGPT 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 (see, for example, <https://apastyle.apa.org/blog/how-to-cite-chatgpt>). Be aware that students are responsible for any errors or information that is misrepresented or inaccurate (i.e., hallucinations) that generative AI tools produce when submitting work that includes AI-generated material.

Students should also include the following statement in their assignments to indicate use of a generative AI tool: *"The author(s) acknowledge the use of [generative AI tool Name] in the preparation or completion of this assignment. The [generative AI tool Name] was used in the following way(s) in this assignment: [e.g., brainstorming, grammatical correction, citation, which portion of the assignment]."*

Important Note on Data Protection and Privacy: When using generative AI tools, be aware that the data you provide may be used for training AI models or other purposes. Consequently, there is no guarantee that the information you provide will remain confidential. You should exercise caution and avoid sharing any sensitive or private information when using these tools. Examples of such information include personally identifiable information, protected health information (PHI), financial data, intellectual property, original research, and any other data that might otherwise be legally protected.

VII. Syllabus Policies, Academic Integrity, Plagiarism

All students are required to read and abide by the Code of Student Academic Integrity. Violations of the Code of Student Academic Integrity, including plagiarism, will result in disciplinary action as provided in the Code. Definitions and examples of plagiarism are set forth in the Code and

on the Student Accountability & Conflict Resolution website. The Code is available from the Dean of Students Office or online at legal.charlotte.edu/policies/up-407. Additional resources are available on the Student Accountability & Conflict Resolution website.

Violation of these syllabus policies may result in appropriate academic penalties, including a reduction of grade in the relevant assignment, project, or exam. If violation of these syllabus policies also implicates the Code of Student Academic Integrity because of alleged academic misconduct, I will follow the process outlined in the Code to address such cases.

VIII. Reporting Expectations

UNC Charlotte is committed to maintaining an environment conducive to learning for all students and a professional workplace for all employees. The University takes active measures to create or restore a respectful, safe, and inclusive environment for community members that is free from discrimination, discriminatory harassment, and interpersonal violence. If you (or someone you know) has experienced any of these incidents, know that you are not alone. UNC Charlotte has staff members trained to support you in navigating campus life, accessing health and counseling services, providing academic and housing accommodations, helping with civil protective orders, and more.

Please be aware that all UNC Charlotte employees, including faculty members, are expected to relay any information or reports of discrimination, discriminatory harassment, or sexual and interpersonal misconduct they receive to the Office of Civil Rights and Title IX. This means that if you tell me about a situation involving these matters, I am expected to report the information. Although I am expected to report the problem, you will still have options about how your case will be handled, including whether or not you wish to pursue a formal complaint. Our goal is to make sure you are aware of the range of options available to you and have access to the resources you need.

If you wish to speak to someone confidentially, you can contact the following on-campus resources, who are not required to report the incident to the Office of Civil Rights and Title IX: (1) Center for Counseling and Psychological Services (CAPS) (caps.charlotte.edu, 7-0311); or (2) Student Health Center (studenthealth.charlotte.edu, 7-7400). Additional information about your options is also available at civilrights.charlotte.edu under the “Students” tab.