Cloud Computing for Data Analysis
ITCS 6190/8190 – Fall 2025

Welcome to *ITCS 6190/8190 – Cloud Computing for Data Analysis*! This is a challenging course, encompassing a substantial amount of 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 learned principles to design and implement data analysis jobs on top of 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.

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.

Location and Time

Tuesday, 11:30 am-2:15 pm, Dubois Center (Uptown) 501

Instructor

Marco Vieira
Email: marco.vieira@charlotte.edu
Office: Woodward 205C
Office Hours:
Tuesday, 10:30 am to 11:30 am, Dubois Center (Uptown) 713
Wednesday, 11:00 am to 12:00 pm, Woodward 205C

Teaching Assistants *(see canvas for details)*

Ajitesh Reddy Tippireddy, atippir1@charlotte.edu
Vinay Devabhaktuni, vdevabh1@charlotte.edu

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, AWS services, and Spark on AWS.

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 11:30 am)**: 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 will complete a short 10-minute quiz, which collectively accounts for 6 points out of the total 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, following 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, documenting their progress through 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 September 30. The final exam, which will account for 20 points, is scheduled for a date yet to be determined. 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%: F

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 three or more days after the deadline will result in zero 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 as well as 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. 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. Classroom Conduct

We will conduct this class in an atmosphere of mutual respect. Active participation in class discussions is greatly encouraged. Each of us may have strongly differing opinions on the various topics of class discussions. The conflict of ideas is encouraged and welcome. The orderly questioning of the ideas of others, including mine, is similarly welcome. However, I will exercise my responsibility to manage the discussions so that ideas and argument can proceed in an orderly fashion. You should expect that if your conduct during class discussions seriously disrupts the atmosphere of mutual respect I expect in this class, you will not be permitted to participate further.

III. Attendance and Absences

Students are expected to attend every class and remain in class for the duration of the session. Failure to attend class or arriving late may impact your ability to achieve course objectives, which

could affect your course grade. An absence, excused or unexcused, does not relieve a student of any course requirement. Regular class attendance is a student's obligation, as is a responsibility for all the work of class meetings, including tests and written tasks.

The instructor has the authority to excuse a student's class absence(s) and to grant a student an academic accommodation (turn in a late assignment, provide extra time on an assignment, reschedule an exam, etc.). However, under Academic Affairs Policy on Course Attendance and Participation, University-sanctioned events or activities are considered excused absences. A University-sanctioned event or activity is one in which a student formally represents the University to external constituencies in athletic or academic activities. This policy does not supersede individual program attendance and/or participation requirements that are aligned with accreditation or licensure. For more information and student responsibilities to account for such an absence, see provost.charlotte.edu/policies-procedures/academic-policies-and-procedures/course-attendance-and-participation.

IV. Instructor's Absence or Tardiness

If I am late in arriving to class, you must wait a full 20 minutes after the start of class before you may leave without being counted absent, or you must follow any written instructions I may give you about my anticipated tardiness.

V. 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.

VI. 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 make an appointment to speak with me as soon as possible. The University policy on withdrawal allows students only a limited number of opportunities available to withdraw from courses. It is important for you to understand the financial and academic consequences that may result from course withdrawal. See: provost.charlotte.edu/policies-procedures/academic-policies-and-procedures/withdrawal-and-cancellation-enrollment-policy

VII. Computer, Cell Phones and Other Mobile Devices in the Classroom

Students are permitted to use computers during class for note-taking and other class-related work only. Those using computers during class for work not related to that class must leave the classroom for the remainder of the class period.

The use of cell phones, smartphones, or other mobile communication devices is disruptive and is therefore prohibited during class. Except in emergencies, those using such devices must leave the classroom for the remainder of the class period.

VIII. Use of Generative Artificial Intelligence (AI)

In this course, students are permitted to use generative artificial intelligence (AI) tools like 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) acknowledges 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, it is important to be aware that the data you supply might 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.

IX. 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.

X. 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.