CS435 Introduction to Big Data

Syllabus

Course Sections: 001 and 801
Semester: Fall 2025
Lecture: 4:00-5:15PM, Monday and Wednesday, Education Building, Room #7
Department of Computer Science, Colorado State University

1. Introduction to Big Data: Course Description

Modern scientific instruments and Internet-scale applications generate voluminous data pertaining to vital signs, weather phenomena, social networks that connect millions of users, and the origins of distant planets. Data produced in these settings hold the promise to significantly advanced knowledge. This course covers fundamental issues in Big Data. The course examines issues related to data organization, storage, retrieval, analysis and knowledge discovery at scale. This will include topics such as large-scale data analysis frameworks, data storage systems, scalable data analytics algorithms and case studies. This course will involve hands-on programming assignments and term project using real-world datasets.

1.1. Course Learning Objectives

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

- Understand fundamental concepts for the big data
- Understand key system's components and scalable algorithms for the big data analysis
- Implement foundational data analysis algorithms using scalable computing frameworks
- Formulate big data analysis problems and apply scalable algorithms alongside comprehensive evaluation
- Design, perform, evaluate, and analyze results of distributed big data analysis over large-scale datasets

1.2. Course Topics

- Overview of Big Data
- Scalable Data Analytics Frameworks
- Large-scale Data Analysis Models (e.g., regression models, random forest, recommendation systems, distributed PageRank algorithm)
- In-memory Data Analytics
- Distributed Storage Architecture
- NoSQL Database

1.3. Course Prerequisites

For undergraduate students:

Undergraduate level Operating Systems (CS370, C or higher) OR undergraduate level Algorithm Theory and Practice (CS320, C or higher) OR undergraduate level Machine Learning Foundations and Practice (CS345, C or higher)

For graduate students: There is no prerequisite for this course.

1.4. Lectures and Help Sessions

Class meeting: Education Room #7 (M, W 4:00PM – 5:15PM MST)

Professor: Sangmi Lee Pallickara, Ph.D.
Email: Sangmi.Pallickara@colostate.edu
Office: CSB 374
Office Hours: Fridays, 9:00–9:50 AM (MST), and by appointment

- Available in-person and online (via MS Teams)
- If this time slot does not work for you, please feel free to request an appointment via email.

Graduate Teaching Assistants
Shivarshini Ksheerasagar
Sofia Catalan
GTA email address: Compsci_CS435@colostate.edu

Recitation classes: 10-10:50AM on Thursdays (Recording will be provided)

Help Desk hours
Hybrid help desk hours are available during regular work hours.

- In-person location: CSB120
- Online: via MS Teams course page

Online-only help desk hours are available after work hours and on weekends.
Please see the Canvas course page for more details.

1.5. Meet Your Professor (MYP) Session for Students in 801 Section

During the first 2–3 weeks of the semester, online students in Section 801 will be invited to join a brief online meeting with the professor. This meet-and-greet is a great opportunity to connect, ask questions, and get to know your instructor. A sign-up sheet will be available on the course Canvas page

1.6. Course Materials

Course Materials
All course materials will be available through Canvas except for interactive online course activities.

Interactive online activities (help sessions, online meetings, and discussions):
Please use the MS Teams course page.

Textbooks:
No textbook required.
Reading lists and online materials are provided through Canvas

2. Course Components and Grading

Course Components	Percentage	Notes
Programming Assignments (Individual)	15%	PA0 (2%) PA1 (3%) PA2 (7%) PA3 (3%)
Term Project (Group)	20%	Team assignment: 1% Proposal: 3% Final report: 10% Final presentation: 3% Team participation: 3% *Note: For the individual scores may be different within a team based on the participation.
Exams	60%	Midterm 30% Final exam: 30%
Quizzes (Individual)	5%	12-14 quizzes (3 lowest scores will be eliminated)
Extra credits	4%	Worksheets (Individual) In-class participations (e.g., students' presentations)

Participation Requirements
This course includes collaborative team activities. Your active participation is required. Your participation will be assessed by your peer team members and accounted for your final grade.

Make Up Policy
Exams: Makeup exams are only given in extraordinary circumstances. If you have a family or medical emergency that prevents you from attending the midterm, or the final, students must consult with the professor as soon as possible, preferably before the start of the exam. Students must provide a valid document to reschedule any course activities. Course examination dates are listed in the syllabus; be aware of them and plan accordingly. No make-ups will be given for missed quizzes.

Letter Grading Policy
Letter grades will be based on the following standard breakpoints: >= 90 is an A, >= 89 is an A-, >=88 is a B+, >=80 is a B, >=79 is a B-, >=78 is a C+, >=70 is a C, >=60 is a D, and <60 is an F. I will not cut higher than this, but I may cut lower.

Late Submission Policy
All of the assignments are to be submitted electronically using Canvas. Always check the assignment page for due dates. Assignments can be submitted up to a maximum of 2 days past the deadline. There will be a deduction of 10% (of the total score for the assignment) as the penalty per day. For example, if the

assignment was due at 5:00 pm on Thursday: (1) you will lose 10% of the total score if you submit between 5:01 pm Thursday and 5:00 pm Friday, (2) if you submit between 5:01 pm Friday and 5:00pm Saturday, you will lose 20% of the points set aside for this assignment. (3) No submissions will be accepted after 5:00pm on Saturday and you will be given a 0 for that assignment.

Flex Tokens for Programming Assignments
You will start the semester with two Flex Tokens. A Flex Token waives the penalty of one day late submission. You can use these tokens either together or separately for two different programming assignments. The flex tokens are applicable *only* for the programming assignments.
If the submission deadline is September 1 and you submit your assignment on September 2, there will be no score deduction with your 1 Flex token. These tokens cannot be reused. If you wish to use your token(s), you must inform the GTA when you submit the assignment. The GTAs will track the number of tokens you use. You are not allowed to apply any partial token.

3. Exams

Midterm
Section 001: 4:00PM--5:15PM, October 8, 2025 in class
Section 801/802: Online exam – a single 75-minute session between 4:00PM October 12, 2025 and 3:59PM October 13, 2025. Students may take the exam any time in this window.

Final exam
Section 001: 11:50AM -- 1:50PM, December 17, 2025 Education Room #7
Section 801/802: Online exam – a single 120-minute session between 11:50AM December 17, 2025 -- 11:49AM December 18, 2025 (Students may take the exam any time in this window. Students are given 120 minutes to complete the exam.)

For the CSU Final exam schedule: [LINK]

4. Weekly Schedule (Tentative, Subject to Minor Adjustment)

Please check the canvas page for the schedule every week.

Week	Topics	Key Deliverables	In-class Activities
1	Introduction Overview of Big Data Introduction to MapReduce	MYP session (online students)	Quiz/Worksheet Survey
2	Introduction to MapReduce MapReduce Design Patterns	MYP session (online students)	Quiz/Worksheet

3	MapReduce Design Patterns		Quiz/Worksheet
4	MapReduce Design Patterns	PA1 submission	Quiz/Worksheet
5	MapReduce Design Patterns		Quiz/Worksheet
6	Web-Scale Link Analysis		Quiz/Worksheet
7	Web-Scale Link Analysis Midterm (10/8/2025)	Midterm	Worksheet, midterm preparation review
8	Web-Scale Link Analysis Clustering algorithms	PA2 submission	Quiz/Worksheet
9	Recommendation systems		Quiz/Worksheet
10	In-memory computing (Apache Spark)	PA3 submission	Quiz/Worksheet
11	Scalable Machine Learning (using Apache Spark)	Term Project Proposal submission (11/4)	Quiz/Worksheet
12	Scalable Machine Learning (using Apache Spark)		Quiz/Worksheet
13	Distributed File Systems		Quiz/Worksheet
14	Fall Break		No Class
15	NoSQL Databases	Final Term Project report and software submission (12/4)	Quiz/Worksheet
16	Term Project Presentations	Final Presentation	Students Presentations
Finals week	Final Exam (12/17)	Final Exam	Final Exam

5. Course Policy

This course will adhere to the CSU Academic Integrity Policy as found on the Students' Responsibilities page [Link] of the CSU General Catalog and in the Student Conduct Code [Link].

5.1. Use of General AI

Use of AI tools such as ChatGPT, Grammarly, Claude and/or their ilk to write or "improve" your code or written work at ANY stage is prohibited. **Turning in code or an essay (e.g., project report) written by generative AI tools will be treated as turning in work created by someone else, namely an act of plagiarism and/or cheating.**

Ultimately, you will get out of the class what you put into it. Simply copying and pasting code or text from ChatGPT/Claude will not lead to any learning. If you are stuck on the assignments, reach out to the course staff.

If any academic plagiarism is detected, the case will be reported to the University, and the record will appear in your diploma and future transcript.

There are multiple reasons why these tools are detrimental to your learning experience:

* They rob you of the ability to think and learn the concepts for yourself since solving problems is an essential step to gaining a solid understanding of the material.

* You will struggle with the in-classroom quizzes and exams where you will not have access to these tools.

* While we acknowledge that these tools are likely to become an important of a software engineer's workflow in the future, you are much more likely to use these tools in an effective manner if you already have expertise in the relevant technical topics. Developing such expertise requires putting in the effort to learn these topics without the assistance of these tools.

5.2. Academic Integrity

We take academic integrity seriously. At minimum, academic integrity means that no one will use another's work as their own. The CSU writing center defines plagiarism this way:

"*Plagiarism is the unauthorized or unacknowledged use of another person's academic or scholarly work. Done on purpose, it is cheating. Done accidentally, it is no less serious. Regardless of how it occurs, plagiarism is a theft of intellectual property and a violation of an ironclad rule demanding credit be given where credit is due.*"

If you plagiarize in your work you could lose credit for the plagiarized submission, fail the assignment, or fail the course. Plagiarism could result in expulsion from the university. Each instance of plagiarism, classroom cheating, and other types of academic dishonesty will be addressed according to the principles published in the CSU General Catalog (see page seven, column two: [Link]).

Of course, academic integrity means more than just avoiding plagiarism. It also involves doing your own reading and studying. It includes regular class attendance, careful consideration of all class materials, and engagement with the class and your fellow students. Academic integrity lies at the core of our common goal: to create an intellectually honest and rigorous community. Because academic integrity, and the personal and social integrity of which academic integrity is an integral part, is so central to our mission as students, teachers, scholars, and citizens, we will ask to you sign the CSU Honor Pledge as part of completing all of our major assignments and assessments.

In this course, every element of class assignments must be fully prepared by the student. The use of generative AI tools for any part of your work will be treated as plagiarism. If you have questions, please contact me.

We will ask each of you to write (or agree) and sign in the following statement on your paper exams or online exam.
"*I have not given, received, or used any unauthorized assistance.*"

5.3. COVID

Important information for students: All students are expected and required to report any COVID-19 symptoms to the university immediately, as well as exposures or positive tests (even home tests). If you suspect you have symptoms, or if you know you have been exposed to a positive person or have tested positive for COVID, you are required to fill out the COVID Reporter.
https://covid.colostate.edu/reporter/

If you know or believe you have been exposed, including living with someone known to be COVID positive, or are symptomatic, it is important for the health of yourself and others that you complete the online COVID Reporter. Do not ask your instructor to report for you.
If you do not have internet access to fill out the online COVID-19 Reporter, please call (970) 491-4600.
You may also report concerns in your academic or living spaces regarding COVID exposures through the COVID Reporter. You will not be penalized in any way for reporting.

When you complete the COVID Reporter for any reason, the CSU Public Health office is notified. Students who report symptoms or a positive antigen test through the COVID Reporter may be directed to get a PCR test through the CSU Health Network's medical services for students.

5.4. Inclusion

We create and nurture inclusive environments and welcome, value and affirm all members of our community, including their various identities, skills, ideas, talents and contributions.

This manifests in this class in our in-class discussions, discussion posts, and our interactions on Teams. You are to be kind to others with various backgrounds and identities, without making assumptions. With that said, this course is considered a 'Brave Space'. This does not exclude the classroom from being a Safe Space - and know you are welcome - and the classroom is meant to challenge ideals. We will often approach topics that may make you feel uncomfortable and challenge your point of view. We will even ask you to argue the opposite of what you believe. That is the joy of learning.

5.5. Copyright of the Course Materials

The course material of this course is the property of the instructor who developed this course and Colorado State University. Posting course materials on external sites (commercial or not) violates both copyright law and the CSU Student Conduct Code. Students who share content without instructor's permission could face legal action. If you maintain your code related to the course activities in a shared code repository (e.g., GitHub), ensure that your code remains private.