

[CS 6220] Syllabus

[Big Data Systems and Analytics]

[Fall 2026]

Instructor Information

Instructor: Ling Liu

Email: ling<DOT>liu<AT>cc<DOT>gatech<DOT>edu

General Course Information

Description

Data has been the No. 1 fast growing phenomenon on the Internet in the last decade. Big data demands both high performance computing and elastic utility driven computing. Big data analytics holds the potential to reveal deep insights such as social influence among customers by analyzing business transactions, user-generated feedback ratings, social and geographical data. In the past 40 years, data was primarily used to record and report business activities and scientific events, and in the next 40 years data will be used also to derive new insights, to influence business decisions and to accelerate scientific discovery.

This course will cover the concepts, techniques, algorithms, and systems of big data systems and data analytics, with strong emphasis on big data processing systems, fundamental models and optimizations for data analytics and machine learning, which are widely deployed in real world big data analytics and applications. The course will navigate the students by examining various research problems and challenges that are critical for developing big data systems and big data applications, as well as exploring big data opportunities from a variety of science and engineering applications and.

Main topics to be covered include but not limited to: fundamentals of big data computing platforms and fundamentals of big data analytic systems and optimizations, incl. emerging trends, such as GenAI, Large Language and vision reasoning, fine-tuning and domain adaptation technologies, big data computing systems and software design, fundamentals of geographically distributed data intensive systems. We will also cover big data applications that pose new challenges to big data systems and analytics, incl. healthcare, mobile commerce, social media, Internet of Things, software defined computing, cyber manufacturing, cyber-physical systems, to name a few.

This course is designed to provide the fundamental training for big data scientists from high-performance Bigdata computing systems, to big data analysis and management algorithms and big data applications. It will also look beyond the present status of the Big Data systems and analytics and conjecture what possible future technologies and applications will evolve. The course will include a significant project component that will typically require programming skills in conventional programming languages.

Course Learning Outcomes

Upon successful completion of this course, the students will learn the concepts, algorithms, optimizations of big data systems, big data technology, big data programming, big data services, and big data applications from both research and development perspectives. Students will gain better understanding of the key challenges and *some* technical solutions that are critical for advanced application developers and software engineers in the broader fields of Big Data systems and analytics. For example, students will learn the key challenges and technologies for handling big data that are challenging in volumes, velocity, and variety (text, audio, image, video, etc.), and learn how to select the right platforms and tools to make reasoning of big data easy and simple. Students will also learn the key challenge and technologies that have and continue to revolutionize the ways of collecting, processing and analyzing the massive data, especially when the big data to be processed and analyzed exceeds the processing capacity of existing computing systems (CPU, GPU, Memory, Network).

Required Course Materials

There is no suitable textbook for this course.

The course materials include lecture notes from the instructor and the required reading list (a collection of papers, one per week).

Grading Policy:

Students will be evaluated by five homework assignments, a group project, and a take-home final exam (technology review).

Grades will be computed using the following weighting scales:

- **Assignments: 25%** [5% for each of the ten assignments]
- **Project: 45%**
- **Final-exam: 25%**
- **Participation: 5%**

Description of Graded Components

Homework Assignment (25%)

There are a total of 5 homework assignments and on average one assignment every 2 weeks. Each assignment requires a student to choose from two types of assignments: reading critique or programming.

For each reading critique assignment, a student should read 2 papers in the same subject area of big data systems and analytics, and write one reading critique for each of the two papers you read. It is recommended that you choose the two papers that are relevant to the lecture topics covered by lectures in the respective two weeks. Each paper critique should be 2 to 5 pages using time Newman font 11 or Arial font 10, single column. Each paper critique should not exceed 5 pages. Each of your two reading critiques should consist of the following sections: (i) overview using your own words, (ii) strong points of the paper, (iii) weak points and limitations of the paper if any, and (iv) your suggestions for improvement (optional).

For each programming assignment, a student should choose one assignment from the list of the programming problems given for each of the two weeks. Most of the programming assignments are designed to help you gain better understanding on certain aspects of big data systems and analytics. Each programming assignment requires the following deliverable: (i) source code, (ii) executable and screen shots of your code execution, (iii) input data structure and 5 examples, (iv) output data structure and 5 examples. and (v) workflow figure on how your code module takes input and product output.

Each of the five homework assignments is graded using the scale of 100 points and is worth 5% of your final grade.

Final Exam/Technology Review (25%)

The final exam is to write a technology review, and it is an open book and take-home exam for the course. Topics should come from weekly lectures, class discussions, guest presentations, as well as homework assignments and topics related to your group project. You are required write a technology review of 15~20 pages with 1-5 pages references in single column single spacing, including figures, using font times Newman size 12 or arial size 11. This technology review paper is due on the final exam day set by GIT in the final exam schedule each semester.

The technology review should cover (1) why this specific technology is of interest to you and your readers, (2) the state of the art research and development to the best of your knowledge, (3) the set of technological aspects that you consider most valuable and important to introduce to your readers, especially some of those are not well known, which you believe are important knowledge about the technology. Here you need to identify your readers are peers or non-CS major or your grandparents or K-12 students. (4) the outlook of the technology in your own words or analysis, including what need to be improved, what new applications can benefit from it, etc. You can create a discussion section dedicated to describing your thought and your prediction of the technology in terms of its impact in the next 10-20 years after surveying this specific technology area. (5) A list of references with citation in your review text.

The technology review topics can be selected based on the weekly lecture theme covered in the course as well as topics in homework assignments. The technology review topic can also be combined with the theme of your course project. You are encouraged to choose a topic that is not too general, such as data mining. Examples of good topics could be trajectory clustering in mobile domain or in healthcare, or optimization techniques in reinforcement learning. The final exam is graded using the scale of 100 points and is worth 25% of your final grade.

Project (45%)

The course project is a team project with 4-5 members per project group given the large size of this class.

The topic of the project should be in the broad areas of Big Data systems or Big Data analytics. The project should demonstrate some innovations in either project design or project engineering and implementation.

In principle, you can propose creative ideas in novel applications, algorithms, system design, implementation, benchmarking and evaluation. For the students who are currently working part time in companies, it is possible to propose a work-related project that promote blue-sky ideas and innovations. Students are encouraged to come up with your own project ideas and are encouraged to discuss your project ideas with instructor. The project is graded using the scale of 100 points and is worth 45% of your final grade.

Note that all course project related material must be non-proprietary, i.e., the instructor and the GTAs will not sign any non-disclosure agreement just to evaluate a project.

Participation (5%)

In this course, students will earn participation grade in several ways. The goal is to encourage each student in the class to attend the lectures, interact with your peers, and contribute to the Ed-discussion forum of the course. All types of participation are graded on quality. The participation is graded using the scale of 100 points and is worth 5% of your final grade.

Letter Grade Policy

The total grade is the sum in percentages of HWs, Project, Final exam (Technology Review), and Participation grade you have earned. Below is the conversion rules for mapping the total percentage in the scale of 100 points to the letter grade:

- [85-100] corresponds to letter grade A
- [75~85) corresponds to letter grade B
- [65~75) corresponds to letter grade C
- [60~65) corresponds to letter grade D
- [0~60) corresponds to letter grade F

Course Policies

Attendance Policy

This is an in-campus course and according to GIT requirement, a GTA will ask students to sign off the attendance sheets in every lecture starting the week 2 of the semester until the last lecture in the week 16.

Students with special circumstances can be excused of absence. Example scenarios include:

- health emergency with doctor's note,
- family emergency with the note from the dean of students,
- off-campus job interviews with the proof of interview schedule
- majors' career fair with the proof of the schedule

Academic Integrity

Students are expected to maintain the highest standards of academic integrity. All work submitted must be original and properly cited. Plagiarism, cheating, or any form of academic dishonesty will result in immediate consequences as outlined in the university's academic integrity policy.

Georgia Tech aims to cultivate a community based on trust, academic integrity, and honor. Students are expected to act according to the highest ethical standards. Review [Georgia Tech's Honor Code](#) and the student [Code of Conduct](#).

Any student suspected of cheating or plagiarism on a quiz, exam, or assignment will be reported to the Office of Student Integrity, who will investigate the incident and identify the appropriate penalty for violations.

Core IMPACTS

[Core IMPACTS](#) is the University System of Georgia's General Education curriculum. If you are teaching a course that counts towards Core IMPACTS, you should include a syllabus statement about the Core area and associated [career competencies](#). [This resource](#) developed by the Center for Excellence in Teaching and Learning and Online Education at Georgia State University includes template syllabus statements for each of the Core IMPACTS areas that you may adapt for your course.

Accommodations for Students with Disabilities

If you are a student with learning needs that require special accommodation, [contact the Office of Disability Services](#) (404-894-2563) as soon as possible to make an appointment

to discuss your special needs and to obtain an accommodations letter. Please also e-mail me as soon as possible in order to set up a time to discuss your learning needs.

Student-Faculty Expectations Agreement (Acceptable student conduct Statement)

At Georgia Tech, we believe that it is important to strive for an atmosphere of mutual respect, acknowledgement, and responsibility between faculty members and the student body. [The Student-Faculty Expectations](#) articulate some basic expectations that you can have of me and that I have of you. In the end, simple respect for knowledge, hard work, and cordial interactions will help build the environment we seek. Therefore, I encourage you to remain committed to the ideals of Georgia Tech while in this class.

Extra Credit Opportunities

This course will provide students opportunities for extra credit of up to 10%, which will be awarded to those students who excel in one or more tasks in the course, ranging from homework, project, final exam, in-class question and answer (Q&A) section, off-class Q&A discussion, and class participation (incl. attendance, Ed-discussion). The extra credit that a student earns will be applied his or her final numerical grade before converting it to the letter grade.

Campus Resources for Students

Undergraduate Student Academic Success Resources:

- Academic Support: Academic Success and Advising (a unit in the Office of Undergraduate Education & Student Success) provides free support for your courses. Students can attend scheduled supplemental review (PLUS) sessions, stop by Drop-In Tutoring, or schedule a one-on-one appointment through Knack. To explore what options work best for you, please visit us online at success.gatech.edu/tutoring, email us at tutoring@gatech.edu, or come see us at Clough Undergraduate Learning Commons, Suite 283.

Graduate Student Academic and Professional Success Resources:

A list of resources for graduate students is given on the [Office of Graduate and Postdoctoral Education](#) website. Specific information for [current graduate students](#) includes

- [Academic Resources](#) such as the Communications Center, Language Institute, Library, Catalog, Registrar, resources for conducting research, Advocacy and

Conflict Resolution resources, and how to manage unexpected situations that may impact your academic performance;

- Student Resources such as Campus Services, Child Care/Family programs, Health & Wellness, Career Services, and the Student Resource Guide; and
- Professional Development such as the programming from the Career Center and other professional development resources and events”

Student Well-Being:

At Georgia Tech, we are concerned about your overall physical, social, and mental well-being. A comprehensive list of wellness related resources has been compiled and maintained by the Office of the Vice President for Student Engagement and Well-being ([student-resource-guide \(gatech.edu\)](http://student-resource-guide.gatech.edu))