

Fall 2024

DS 644-101, 1J1, 851, 853: Introduction to Big Data

Yajuan Li

Follow this and additional works at: <https://digitalcommons.njit.edu/ds-syllabi>

Recommended Citation

Li, Yajuan, "DS 644-101, 1J1, 851, 853: Introduction to Big Data" (2024). *Data Science Syllabi*. 23.
<https://digitalcommons.njit.edu/ds-syllabi/23>

This Syllabus is brought to you for free and open access by the NJIT Syllabi at Digital Commons @ NJIT. It has been accepted for inclusion in Data Science Syllabi by an authorized administrator of Digital Commons @ NJIT. For more information, please contact digitalcommons@njit.edu.

Course Syllabus - Fall 2024

DS 644: Introduction to Big Data

Instructor: Dr. Yajuan Li

Email: yajuan.li@njit.edu

Office: GITC 4309

Office Hours: Monday 11:40-1:10 pm (Office), 1:40-3:10 pm (Zoom)

Course Overview and Learning Outcomes:

This course provides an in-depth coverage of various topics in big data from data generation, storage, management, transfer, to analytics, with focus on the state-of-the-art technologies, tools, architectures, and systems that constitute big-data computing solutions in high-performance networks. Real-life big-data applications and workflows in various domains (particularly in the sciences) are introduced as use cases to illustrate the development, deployment, and execution of a wide spectrum of emerging big-data solutions. When you have completed this course, you should be familiar with big data tools, techniques, and systems, and be able to analyze/solve big data problems.

Required Background:

Programming Skills

- Java, Python, or C/C++ in Linux

Prerequisite Courses

- CS 610: Data Structures and Algorithms
- Or permission of instructor

Textbook (optional): Big Data Analytics: From Strategic Planning to Enterprise Integration with Tools, Techniques, NoSQL, and Graph. By David Loshin, Elsevier, August 23, 2013.

Attendance: You are supposed to attend all the classes. Participation is highly encouraged to make the class more interactive. In general, students who attend class regularly perform much better than those who come only occasionally. If you miss one class, be sure to consult one of your classmates about the content of the lecture and use canvas to get notes, exercises, assignments, deadlines, and announcements.

Late Policies

(1) Late submissions for homework/project will be accepted, but will be penalized.

Days Late (x)	Penalty (percentage off the full grade)
$0 < x \leq 1$	25%
$1 < x \leq 2$	50%
$2 < x \leq 3$	75%
$x > 3$	100% (will not be accepted)

(2) Students can get extended deadline if they have special/emergency reasons verified from Dean of students. <https://www.njit.edu/dos/student-excuses>

There will be NO EXCEPTION to these late policies. Please manage your time appropriately.

Exam Policies

There will be one midterm and one final exam. Be sure that you will be present for all of your exams. Respondus LockDown Browser and Monitor can be used to monitor and proctor the test. All students are required to have a working webcam in order to take the exam. More information about Respondus Lockdown Browser can be found at

<https://web.respondus.com/student-help/>

You must bring a student ID to all exams. There are no late submission or makeup for exams. Students who have special/emergency reasons to reschedule exams must apply and get approval of testing accommodation from the Office of Accessibility Resources and Services before the exams. <https://www.njit.edu/accessibility/requesting-testing-accommodations>

Course Grade:

- Homework (15%)
- Project (20%)
- Midterm Exam (30%)
- Final Exam (35%)

Grading Scale:

Grade	A	B+	B	C+	C	F
Overall Course Score	≥ 93	[86, 93)	[78, 86)	[70, 78)	[60, 70)	< 60

*Final grade will not be curved unless necessary.

Tentative Course Topics (Subject to changes according to progress)

Week	Topic
1	Introduction
2	In-class presentation on 4v's of big data applications
3	Computing trends for big data
4,5	Big data overview
6~8	Big data tools, techniques, and systems
9	Midterm exam
10~12	Big data analytics (clustering, classification, regression, etc)
13	Big data visualization
14	Review
15	Final exam

Collaboration and Honor Code

Each student is responsible for his/her own work. Students may discuss problems together but must write up their own solutions. When writing up the solutions, students should write the names of people, if any, with whom they discussed the assignment. Note that copying homework or programming assignments, in full or in part is forbidden. Students found cheating or plagiarizing will be immediately referred to the Dean of Students and the NJIT Committee on Professional Conduct and subject to Disciplinary Probation, a permanent marking on the record, possible dismissal, and an "F" grade in the course. All submitted assignments will be checked for similarities, and plagiarism and guilty students identified. In the exam, each student is required to sign the Honor Code Agreement "On my honor, I pledge that I have not violated the provision of the NJIT Student Honor Code."

University Policy on Academic Integrity

"Academic Integrity is the cornerstone of higher education and is central to the ideals of this course and the university. Cheating is strictly prohibited and devalues the degree that you are working on. As a member of the NJIT community, it is your responsibility to protect your educational investment by knowing and following the academic code of integrity policy that is found

at: <http://www5.njit.edu/policies/sites/policies/files/academic-integrity-code.pdf>.

Please note that it is my professional obligation and responsibility to report any academic misconduct to the Dean of Students Office. Any student found in violation of the code by cheating, plagiarizing or using any online software inappropriately will result in disciplinary action. This may include a failing grade of F, and/or suspension or dismissal from the university. If you have any questions about the code of Academic Integrity, please contact the Dean of Students Office at dos@njit.edu"

Generative AI

Generative AI tools, such as ChatGPT, are permitted in this course only if explicitly authorized by the instructor for a specific assignment. The student is fully responsible for the integrity and accuracy of all submitted work. If permitted Generative AI tools produce incorrect results that are incorporated into the submission, the student will be graded based on the accuracy of the submitted work. Errors made by AI tools are solely the responsibility of the student.

*Students will be notified in class of any changes to the syllabus.