CSCI/DASC 6010: Big Data Analytics and Management

Spring 2025: Jan 13 – May 8
3 Credits

Contact me ✉	Meet with me	Class location and time	Course materials
Nic Herndon herndonn19@ecu.edu	TTh 9:30 – 11:00 & 3:30 – 4:30 SciTech C-108 and MS Teams	TTh 11:00 – 12:15 PM Howell N107 & **MS Teams**	Listed in Section 1.3

Contents

1 Course description

Approaches to storing, processing, retrieving, analyzing, and managing massive-scale structured and unstructured data. High-performance computing architectures and methods for developing and querying databases for Big Data. Column-relational, key-value, column-oriented, RDF, document-oriented, native XML, and graph databases.

1.1 Course objective

This course will introduce methods and applications for managing and analyzing big data – any collection of data sets so large or complex that it becomes difficult to process them using traditional data management techniques. During the course students will demonstrate an understanding of the fundamental principles underlying

programming languages and demonstrate an ability to design and implement applications in a variety of language paradigms.

By the end of the semester students will:

- Demonstrate an understanding of the typical problems encountered when scaling a traditional database.

- Understand the strengths and weaknesses of NoSQL and graphical databases.

- Be able to explain the desired properties of a big data system.

- Design and implement an analysis workflow to analyze big data.

- Exercise public speaking skills by discussing various course topics with classmates, asking and answering questions, and doing group presentations.

By the end of this course students should be able to transfer the knowledge gained, and apply it outside of the context of the course to:

- Identify, formulate, analyze, and solve complex computing or engineering problems by applying principles of computing, engineering, science, and mathematics.

- Design, implement, and evaluate a computing or engineering solution to meet a given set of requirements, with consideration of public health, safety, and welfare, as well as global, cultural, social, environmental, and economic factors.

- Communicate effectively in a variety of professional contexts, with a range of audiences.

- Function effectively as a member or leader of a team engaged in activities appropriate to the program's discipline, creating a collaborative and inclusive environment, establishing goals, planning tasks, and meeting objectives.

1.2   Prerequisites

Enrolled in the master of science in computer science, software engineering, or data science programs or consent of instructor.

1.3   Optional course materials

- *Big Data: Principles and best practices of scalable realtime data systems*

- *Introducing Data Science: Big data, machine learning, and more, using Python tools*

- *Data Science at the Command Line*

- *PySpark* and *Scala Spark programming guide*

1.4   Tentative schedule

Lecture	Day	Date	Topic
1	Tue	01/14	Introduction, final project, and big data tools
			- Pitch project idea (3 min), due Sun 2/2
			- Big data tool (homework 6), due Sun 3/23
2	Thu	01/16	Vector databases
			- Homework 1, due Sun 1/26
3	Tue	01/21	Vector databases
4	Thu	01/23	Data wrangling with Linux power tools, OpenRefine
			- Homework 2, due Sun 2/2
5	Tue	01/28	Public speaking (recorded)
6	Thu	01/30	Relational databases
			- Homework 3, due Sun 2/9
7	Tue	02/04	Project ideas (recorded)
			- Form teams by Sun 2/16, submit proposals by Sun 3/2
8	Thu	02/06	Project ideas (recorded)
9	Tue	02/11	Relational databases
10	Thu	02/13	Graph databases
			- Homework 4, due Sun 2/23
11	Tue	02/18	Graph databases
12	Thu	02/20	Graph databases
13	Tue	02/25	Data visualization
14	Thu	02/27	Machine learning + AutoML
15	Tue	03/04	Project proposals (recorded)
16	Thu	03/06	Project proposals (recorded)
	Tue	03/11	Spring Break 3/9 – 3/16
	Thu	03/13	
17	Tue	03/18	Document databases
			- Homework 5, due Sun 4/6
18	Thu	03/20	Document databases
19	Tue	03/25	Big data tools (recorded)
20	Thu	03/27	Big data tools (recorded)
21	Tue	04/01	Big data tools (recorded)
22	Thu	04/03	Document databases
23	Tue	04/08	Apache Spark
			- Homework 7, due Sun 4/20
24	Thu	04/10	Apache Spark
25	Tue	04/15	Apache Spark
26	Thu	04/17	Final project presentations (recorded)
27	Tue	04/22	Final project presentations (recorded)
28	Thu	04/24	Final project presentations (recorded)

2 Course staff

My name is Nic Herndon. I worked in industry and academia, in both public and private sectors, and this is the sixth time I teach this course. Assisting me this semester are Simon Polishchuk (polishchuks22@students.ecu.edu) and Kehinde Akinola (akinolak22@students.ecu.edu). If you need to communicate with us outside of class and student hours, please do so via email. We'll do our best to respond within two business days.

3 How to succeed in this course

Your success in this class is important to me. We all learn differently and bring different strengths and needs to the class. If there are aspects of the course that prevent you from learning or make you feel excluded, please let me know as soon as possible. Together we'll develop strategies to meet both your needs and the requirements of the course. There are also a range of resources on campus, including:

- Center for Communication Excellence – can help with managing speaking anxiety, working on a professional social media presence, developing professional interpersonal communication skills, and developing engaging public speaking skills. Do you wish you felt more confident when communicating with your professors and supervisors? Do you dread getting up to speak in front of your peers? Would you like to be a more engaging speaker when giving class presentations? Do your professional communication skills need help? The Center for Communication Excellence, located in 205 Joyner East, is available to you as a FREE service. Face to face and virtual appointments are available. For more information and hours of operation, call 252.328.2790 or email commcenter@ecu.edu.

- Computer Science Peer Tutoring – peer tutoring for a number of undergraduate courses in Computer Science and Software Engineering. The tutoring schedule for the current semester can be found online.

- Pirate Academic Success Center – the center offers free tutoring for science and humanities courses, a variety of math and business courses, foreign languages, and study skills such as time management, test taking, and effective textbook reading. Call (252) 737-3009, email tutoring@ecu.edu or visit 2300 Old Cafeteria Complex to make an appointment. All services are free to ECU students.

- Center for Counseling and Student Development – provides individual counseling within a short-term therapeutic model, with most students attending 4-5 sessions.

- ECU CARES – offers assistance to distressed individuals, connecting them to appropriate campus resources, and reporting concerning behavior to professionals on campus. This can be a critical first step in helping the individual(s) improve and ensuring a safer campus for everyone.

- Language Academy – a premier destination for students interested in improving their English skills and understanding of American culture.

3.1 Feedback

I encourage every student to submit anonymous feedback throughout the semester on how to improve this course. In addition, during the lectures we'll be using this etherpad which can be used to post anonymous questions and to chat with classmates.

3.2 Equality statement

I am dedicated to establishing a learning environment that promotes diversity of the students including race, class, culture, religion, gender, sexual identity, and physical ability. It is important that this is a safe classroom environment. We will practice being generous and respectful members of our class and computer science community. Please let me know immediately if you notice discriminatory behavior in this class or feel discriminated against.

3.3 Accommodations

East Carolina University seeks to comply fully with the Americans with Disabilities Act (ADA). Reasonable accommodations will be made for students with verifiable disabilities. In order to take advantage of available accommodations, students must be registered with the Department for Disability Support Services located in Slay 138, 252-737-1016. Additional DSS student resources can be found at: https://accessibility.ecu.edu/students/.

3.4 Student hours

Student hours are an important part in supporting you throughout this course. Even if you don't have specific questions, needs, and concerns, I would love to meet up with you at least once during this semester. There are a couple of options to meet up: meet with me using MS Teams during the student hours, or, if these times don't work for you, please email me and I will find another day and/or time.

3.5 Attendance and participation

- Face-to-face students are expected to attend, and be on time, for every class. This demonstrates professionalism and consideration for your fellow students and your instructor. While the course does not have an attendance policy, students who miss class and/or are late for class may experience an impact on their grade by missing classroom activities.

- Distance education students are expected to demonstrate professionalism and courtesy by muting their microphones if they attend the lectures streamed live, and unmute them when participating in class activities (Q&A, ICEs, etc.).

- For group work, students are expected to clearly state to their teammates, their expectations on communications (in person, email, Microsoft Teams, etc.), deadlines, and contributions of each student. If the agreed expectations are not met by one of the students, I encourage the teammates to reach out to him/her, and if a resolution cannot be reached, to inform me right-away.

- Students should make every effort to participate in class activities.

- Missing class does not automatically result in extensions on assignments or exam due dates.

- There is a clear correlation between attendance/participation and your grades.

- It is the student's responsibility to seek out the instructor and other classmates to obtain the information (e.g., lecture notes, relevant announcements, etc.) if they missed class.

- Students should maintain regular communication with instructors regarding their health status and communicate any changes in their ability to complete coursework and academic responsibilities immediately.

- You may at any time consult with your advisor or the instructor about whether to request an Incomplete (ECU Faculty Manual IV.VIII.E Grade of Incomplete) or a Withdrawal (ECU's Withdrawals Policies).

3.6 Missed instructional time in the event of a disruption

Making up missed instructional time in this course will follow ECU's Policy for Making Up Missed Instructional Time Due to Suspension of Instruction.

3.7 Course staff responsibilities

You can expect from me:

1. Timely release of course assignments.

2. Timely release of scores achieved on course assignments.

3. To respond to questions in the discussion forums in a reasonable amount of time.

4. Be respectful of your ideas and value the diversity you bring to the class.

5. Be open to dialogue that challenges me.

6. Be available during the stated student/office hours.

7. Ensure the proper running of the course.

4 Grading and course assignments

Homework (50%): There will be seven homework assignments (including one presentation/tutorial on big data tools), on topics covered in class. Everyone will get a total of three late calendar days to use for homework assignments. After all three late days have been exhausted, no more late submissions will be accepted. For unforeseen health and personal emergencies, please contact me. Job interviews/other schoolwork are not excuses for late homework. The homework assignments will be posted on the course page in Canvas, and have to be submitted via Canvas. They are due on the dates listed in Canvas by 11:59 PM.

Final project (40%): There will be several assignments as part of the final project, including but not limited to, discussions, teamwork, project proposal presentation, and final project presentation.

Participation (10%): There will be activities in most lectures during which students will have to participate. For on-campus students these will be through Poll Everywhere responses, Etherpad discussions, Canvas discussions, and quizzes. For online students these will be through Etherpad discussions, Canvas discussions, and quizzes (online students can request in writing, via email, to switch to activities for on-campus students, if they wish to do so).

4.1 Regrade requests

If you feel you deserved a better grade on an assignment, you may submit a regrade request via email within three calendar days after the grades are released. Your request should briefly summarize why you feel the original grade was unfair, and indicate which portion of the assignment was graded unfairly. This is not an opportunity to submit an updated assignment, but rather to correct a mistake done during the grading of your assignment. Your TA will reevaluate your assignment as soon as possible, and then issue a decision. If you are still not happy, you can ask for your assignment to be regraded by an instructor. If no request was received within three days, the grade remains final for that assignment.

5 Final grades

To evaluate your understanding of the course content I will use scores achieved on each of the above assessment components. Your final grade will convey what you know from the course and how well you know it. Missing and late assignments can have a dramatic impact on your final grade so it is important that you are attentive to submission deadlines and avoid any missing work. The typical breakdown of percentages and final grades for this course are as follows.

F	C	B	A
$0 - 69$	$70 - 79$	$80 - 89$	$90 - 100$

This grading scheme may be adjusted based on the overall performance of students in the course.

5.1 FAQ regarding final grades

1. Will there be extra credit offered at the end of the semester?
   No. There will be extra credit offered throughout the semester – please take advantage of it.

2. But I had difficulty during the semester due to _____. Isn't there something that can be done?
   Please notify me right away if you have any emergencies; don't wait until the end of the semester to bring this to my attention.

3. My final grade is 89.99%. Will I get an A?
   No. The grade ranges are listed above, and there will be no rounding up. 89.99% earns you a B.

4. I turned in my homework late by 1 minute because the Internet was slow. Will that use any of my late days?
   Yes. Late days are used in one day increments – turning in a homework between 1 second and 24 hours late counts as one late day. Each homework assignment has unlimited submission attempts; so, turn it in early, and turn it in often, just as you would check your code in a `git` repository. We will grade your latest homework submission.

5. I used my three late days for homework 1. Can I use three late days for homework 2?
   No. Each student has a total of three late days for the whole semester. Once you used them you have to turn in on time the remaining homework assignments.

6. Is it okay to work alone without forming a team?
   No. In a professional setting you never work alone (you might work by yourself, but you're a part of the team). Learning how to work well in a team setting is a must.

6 Course incompletes

Students who are unable to complete course requirements within the allotted time because of severe medical or personal problems may request a grade of Incomplete from the instructor of the course. Incomplete grades are warranted only if a student is passing the course at the time of the request and if the course requirements can be completed by the end of the following semester.

Note: an incomplete means you are on your own to complete the material agreed upon by the instructor of this course. Do not expect additional help or one-on-one teaching of the material past the course completion date. It is your responsibility to complete the remaining material.

7 Academic honesty

It is very important in all courses that you are honest in all the work that you complete. You may discuss assignments with other students, in fact I encourage this as a learning experience. But again, the writeup must be your work. Copying is not allowed, and collaboration so close that it looks like copying, is not allowed. Remember to tell me who you worked with as well.

If you copy on project assignments or exams you are doing a disservice to yourself, the instructor for the course, the Department of Computer Science, the East Carolina University, and your future. We design our courses to provide you the necessary understanding and skill that will make you an excellent computer scientist. Assignments are designed to apply and test your knowledge and understanding of the material.

I will carefully review your submissions automatically and manually to verify that "cheating" has not taken place. If you are suspected of plagiarism, I will follow an informal path to determine if academic dishonesty has taken place, and you may receive an F for the course and have a mark on your permanent record at ECU. This will disrupt your schedule for completing courses and may lead to you not completing your degree in a timely fashion. Please review carefully the Academic Integrity to understand what academic dishonesty is, how you can avoid it, and the procedure I will follow if you are under suspicion. If you have questions or unsure if something constitutes plagiarism, please reach out to me.

Every assignment implicitly includes a "contract" that you sign virtually by submitting your assignment. By "signing" this contract you indicate that you have read all the documents on this website and any links to

academic honesty associated with the university. Your signature indicates that you completely understand the policies in place and that you have not plagiarized.

8 Class recordings

This class will be recorded and broadcast on the internet and/or distributed on other electronic media now or hereafter known. These recordings may contain your image and your voice. You must notify me as soon as possible if you DO NOT want your image and your voice contained on the recording.

9 Final note

Occasionally, it may be necessary to revise this syllabus due to extenuating circumstances. I reserve the right to revise this syllabus if the need arises. If I do so, I will provide you with advance notice.