



UMBC

Data 603: Platforms for Big Data Processing

Instructor: Andy Enkeboll

Email: enkeboll@umbc.edu

Spring 2022

Class Number: 7511

Course Description

The goal of this course is to introduce methods, technologies, and computing platforms for performing data analysis at scale. Topics include the theory and techniques for data acquisition, cleansing, aggregation, management of large heterogeneous data collections, processing, information and knowledge extraction. Students are introduced to map-reduce, streaming, and external memory algorithms and implementations using Hadoop and its ecosystem (MapReduce, HBase, Spark, and others). Students are also introduced to modern big data warehousing tools such as BigQuery, Snowflake, Databricks, and Singlestore. The focus of this course is to gain real-world, industry experience in analyzing large datasets and interacting with big data databases.

Prerequisite: Enrollment in the Data Science program. Other students may be admitted with instructor permission. Programming experience.

Corequisite: DATA 601: Introduction to Data Science

Learning Objectives

Upon completion, students will:

- Have the ability to analyze large datasets using modern tools
- Be able to analyze big data architectures for data science applications
- Be familiar with how to use the public domain big data architectures such as Hadoop and its ecosystem
- Know the fundamental concepts, strategies, and pitfalls of processing data at scale
- Be confident in navigating any modern big data warehouse

Meeting Times

Thursday – 7:10 - 9:40, classes via Webex (links to be distributed)

01/31/2022 - 05/17/2022

For Spring 2022 Academic Calendar, refer to <https://registrar.umbc.edu/spring-graduate-academic-calendar/>

Syllabus (Subject to change)

Date	Topic
Week 1 2022-02-03	<ul style="list-style-type: none">• Big Data and Distributed Computing Introduction
Week 2 2022-02-10	<ul style="list-style-type: none">• Distributed File Systems and File Formats• Distributed Databases: HBase/Accumulo, Cassandra
Week 3 2022-02-17	<ul style="list-style-type: none">• Map Reduce Design Patterns• Introduction to Cloud Computing
Week 4 2022-02-24	<ul style="list-style-type: none">• Introduction to Apache Spark• Spark Resilient Distributed Dataset (RDD), Data Frame
Week 5 2022-03-03	<ul style="list-style-type: none">• Spark Structured API Part 1
Week 6 2022-03-10	<ul style="list-style-type: none">• Spark Structured API Part 2
Week 7 2022-03-17	<ul style="list-style-type: none">• Structured Streaming Part 1
Week 8 2022-03-24	<ul style="list-style-type: none">• Structured Streaming Part 2
Week 9 2022-03-31	<ul style="list-style-type: none">• No Class – Spring Break
Week 10 2022-04-07	<ul style="list-style-type: none">• Advanced Analytics Part 1
Week 11 2022-04-14	<ul style="list-style-type: none">• Advanced Analytics Part 2• <i>Project progress report</i>
Week 12 2022-04-21	<ul style="list-style-type: none">• Graph Processing
Week 13 2022-04-28	<ul style="list-style-type: none">• <i>Technical paper due</i>• Monitoring, Debugging, Performance Tuning
Week 14 2022-05-05	<ul style="list-style-type: none">• Project Presentations 1
Week 15 2022-05-12	<ul style="list-style-type: none">• <i>Project presentation due</i>• Project Presentations 2

Materials

Suggested Readings:

- Chambers, Bill, and Matei Zaharia. [*Spark: The Definitive Guide: Big Data Processing Made Simple*](#). O'Reilly Media, 2018.
- Damji, Jules S., et al. Learning Spark: Lightning-Fast Data Analytics 2nd Edition O'Reilly Media, 2018.
NOTE: Free download at: <https://databricks.com/p/ebook/learning-spark-from-oreilly>, use your UMBC email address.

Class Structure (Tentative)

Time	Topic
7:10	Review Homework / Discussion
7:30	Primary Lecture
8:45	Break ~15 Minutes
9:00	Secondary Lecture / Demos
9:30	Review / Quiz / Upcoming
9:40	Adjourn

Format and Assignments

Students will complete 4-7 homework assignments, online and in-class discussions, a semester-long project, and a final paper. Homework will give students an opportunity to gain practical insights with specific big data methods. The project will give students the opportunity to practice big data management by developing a big data processing application using a common big data platform.

Grading Criteria

Course Work	Grade Weight
Homework	25%
Participation	10%
Quizzes	10%
Project	35%
Technical Research Paper	20%

Students are expected to participate in class and online discussions. You will get more out of the class by contributing to it.

Final grade will be computed as follows:

Grade Range	Letter Grade
>93%	A
90-92%	A-
87-89%	B+
83-86%	B
80-82%	B-
77-79%	C+
73-76%	C
70-72%	C-
67-69%	D+
60-66%	D
< 60%	F

Policies

UMBC provides a range of writing assistance, which can be found in the following:

- The Writing Center <http://lrc.umbc.edu/tutor/writing-center/>
- Research Guides & Tutorials <http://lib.guides.umbc.edu/tutorial>
- Failure to follow guidelines for each assignment, including the required format, style, length, submission, etc., may result in at least one-letter-grade reduction on the paper depending on the type and/or number of transgressions.
- Late/Incomplete assignments will not be accepted unless an extension has been agreed to in advance. Emergency situations will be handled on a case-by-case basis with appropriate justification and/or documentation.
- Incomplete grades will not be entertained unless extenuating circumstances warrant and your request is made before the last week of class.

Academic Integrity

By enrolling in this course, each student assumes the responsibilities of an active participant in UMBC's scholarly community in which everyone's academic work and behavior are held to the highest standards of honesty. Cheating, fabrication, plagiarism, and helping others to commit these acts are all forms of academic dishonesty, and they are wrong. Academic misconduct could result in disciplinary action that may include, but is not limited to, failure, suspension, or dismissal.

Refer to the UMBC policy at:

<http://catalog.umbc.edu/content.php?catoid=14&navoid=718#academic-integrity>

Accessibility and Disability Accommodations, Guidance, and Resources

Accommodations for students with disabilities are provided for all students with a qualified disability under the Americans with Disabilities Act (ADA & ADAAA) and Section 504 of the Rehabilitation Act who request and are eligible for accommodations. The Office of Student Disability Services (SDS) is the UMBC department designated to coordinate accommodations that creates equal access for students when barriers to participation exist in University courses, programs, or activities.

If you have a documented disability and need to request academic accommodations in your courses, please refer to the SDS website at sds.umbc.edu for registration information and office procedures.

SDS email: disAbility@umbc.edu

SDS phone: (410) 455-2459

If you will be using SDS approved accommodations in this class, please contact the instructor to discuss implementation of the accommodations. During remote instruction requirements due to COVID, communication and flexibility will be essential for success.

Sexual Assault, Sexual Harassment, and Gender Based Violence and Discrimination

[UMBC Policy](#) and Federal law (Title IX) prohibit discrimination and harassment on the basis of sex, sexual orientation, and gender identity in University programs and activities. Any student who is impacted by sexual harassment, sexual assault, domestic violence, dating violence, stalking, sexual exploitation, gender discrimination, pregnancy discrimination, gender-based harassment or retaliation should contact the [University's Title IX Coordinator](#) to make a report and/or access support and resources:

Mikhel A. Kushner, Title IX Coordinator (she/they)

410-455-1250 (direct line), kushner@umbc.edu

You can access support and resources even if you do not want to take any further action. You will not be forced to file a formal complaint or police report. Please be aware that the University may take action on its own if essential to protect the safety of the community.

If you are interested in or thinking about making a report, please use the [Online Reporting/Referral Form](#). Please note that, if you report anonymously, the University's ability to respond will be limited.

Notice that Faculty are Responsible Employees with Mandatory Reporting Obligations:

All faculty members are considered Responsible Employees, per [UMBC's Policy on Sexual Misconduct, Sexual Harassment, and Gender Discrimination](#). Faculty are therefore required to report any/ all available information regarding conduct falling under the Policy and violations of the Policy to the Title IX Coordinator, even if a student discloses an experience that occurred before attending UMBC and/or an incident that only involves people not affiliated with UMBC. Reports are required regardless of the amount of detail provided and even in instances where support has already been offered or received.

While faculty members want to encourage you to share information related to your life experiences through discussion and written work, students should understand that faculty are

required to report past and present sexual assault, domestic and interpersonal violence, stalking, and gender discrimination that is shared with them to the Title IX Coordinator so that the University can inform students of their [rights, resources, and support](#). While you are encouraged to do so, you are not obligated to respond to outreach conducted as a result of a report to the Title IX Coordinator.

If you need to speak with someone in confidence, who does not have an obligation to report to the Title IX Coordinator, UMBC has several [Confidential Resources](#) available to support you:

- [The Counseling Center](#) (Main Campus): 410-455-2472 / After-Hours 410-455-3230 [Monday – Friday; 8:30 a.m. – 5 p.m.]
- [Center for Counseling and Consultation](#) (Shady Grove Campus): 301-738-6273 (Messages checked hourly)
[Online Appointment Request Form](#)
- [University Health Services](#): 410-455-2542 [Monday – Friday 8:30 a.m. – 5 p.m.]
- Pastoral Counseling via [Interfaith Center](#): 410-455-3657; interfaith@umbc.edu [7 days a week; Fall and Spring 7 a.m. – 11 p.m.; Summer and Winter 8 a.m. – 8 p.m.]

Other Resources:

- Women’s Center (for students of all genders): 410-455-2714; womenscenter@umbc.edu. [Monday – Thursday 9:30am-6pm and Friday 9:30am-4pm]
- [Shady Grove Student Resources, Maryland Resources, National Resources](#).

Child Abuse and Neglect:

Please note that Maryland law and [UMBC policy](#) require that faculty report all disclosures or suspicions of child abuse or neglect to the Department of Social Services and/or the police.

Pregnancy

UMBC’s [Policy on Sexual Misconduct, Sexual Harassment and Gender Discrimination](#) expressly prohibits all forms of Discrimination and Harassment on the basis of sex, including pregnancy. [Resources for pregnant, parenting and breastfeeding students](#) are available through the University’s Office of Equity and Inclusion. Pregnant and parenting students are encouraged to contact the Title IX Coordinator to discuss plans and ensure ongoing access to their academic program with respect to a leave of absence or return following leave related to pregnancy, delivery, adoption, breastfeeding and/or the early months of parenting.

Pregnant students and students in the early months of parenting may be entitled to accommodations under Title IX through the Office of Equity and Inclusion.

In addition, students who are pregnant and have an impairment related to their pregnancy that qualifies as disability under the ADA may be entitled to accommodations through the [Student Disability Service Office](#).

Religious Observations

UMBC [Policy](#) provides that students should not be penalized because of observances of their religious beliefs, and that students shall be given an opportunity, whenever feasible, to make up within a reasonable time any academic assignment that is missed due to individual participation in religious observances. It is the responsibility of the student to inform the instructor of any intended absences or requested modifications for religious observances in advance, and as early as possible. For questions or guidance regarding religious observance accommodations please contact the [Office of Equity and Inclusion](#) at oei@umbc.edu.

Hate, Bias, and Harassment

UMBC values safety, cultural and ethnic diversity, social responsibility, lifelong learning, equity, and civic engagement.

Consistent with these principles, [UMBC Policy](#) prohibits discrimination and harassment in its educational programs and activities or with respect to employment terms and conditions based on race, creed, color, religion, sex, gender, pregnancy, ancestry, age, gender identity or expression, national origin, veterans status, marital status, sexual orientation, physical or mental disability, or genetic information.

Students (and faculty and staff) who experience discrimination, harassment, hate or bias or who have such matters reported to them should use the [online reporting/referral form](#) to report discrimination, hate or bias incidents. You may report incidents that happen to you anonymously. Please note that, if you report anonymously, the University's ability to respond will be limited.