

CSC 343-01, Big Data Programming and Management
Spring 2019
Eastern Connecticut State University

Instructor: Dr. Garrett Dancik
dancikg@easternct.edu
(860) 465-4587
Science Building, Rm 257

Office Hours: MWF: 10-11:00
MW: 3-4:00, or by appointment

Course information:

Title: Big Data Programming and Management
Day/Time: MWF 11:00 – 11:50 AM (SCI 135)
Section: 01
Credit: 3 hours

Course Materials:

1. Course notes and class website: <https://gdancik.github.io>
2. Docker (<http://www.docker.com>) containers will be used to run various Cloudera CDH/Hadoop (<https://hub.docker.com/r/cloudera/quickstart/>) tools
3. Python (<http://www.python.org>) will be used for some programming assignments.
4. Piazza (<https://piazza.com>) will be used for online discussion and several assignments. Note: a mobile app is available from the App store (iPhone/iPad) or Google Play (Android devices)

Course Description

The analysis of Big Data – datasets that are too large or complex to be handled using traditional data storage and processing tools – drives discovery and innovation in science and industry. This course introduces students to the challenges posed by Big Data, along with the theory, systems, and algorithms used to manage and process large data sets in a parallel and distributed fashion. Students are exposed to theoretical and applied topics in parallel and distributed processing, data models, scalable computing and programming models, and levels of abstraction for Big Data processing. The course discusses specific platforms and teaches students how to effectively design and develop Big Data applications.

Grading

Labs / Exercises	20%
Exam I	20%
Exam II	20%
Exam III	20%
Final Project	20%

Online discussion: We will use Piazza (<https://piazza.com>) as an online discussion and question and answer forum in this course. Shortly after the beginning of the semester, you will receive an e-mail with registration instructions sent to your Eastern e-mail address. Piazza allows for students to post and answer questions, anonymously if desired. The class benefits by seeing questions asked by other students (who often have the same questions as you) and by contributing answers. As the instructor, I will answer questions and can endorse correct student answers as well. For these reasons, all non-personal (e.g., not grade-related) questions should be posted to Piazza rather than e-mailed to me. Questions regarding homework assignments should be posted to Piazza. Questions regarding homework must be specific and may contain no more than *several* lines of code. Note that posts not meeting these criteria will be deleted and the poster penalized if warranted. Note that piazza will be required for several assignments.

Exam Policy: Make-up exams will only be given if you have an official excuse for missing class. If you know ahead of time that you will miss an exam, please talk to me before the exam to make arrangements for taking it. Missing **two** or **more** exams without official excuses will result in your dismissal from the course with a grade of **F**.

Grading Scale

93-100: A	90-92: A-	
87-89: B+	83-86: B	80-82: B-
77-79: C+	73-76: C	70-72: C-
65-69: D+	60-64: D	59 and below: F

Academic Honesty

You are encouraged to discuss projects and exercises with one another unless specified otherwise. However, copying answers from another student (unless otherwise specified) is *cheating* and this will not be tolerated. A student found cheating will automatically receive a grade of "F" on the assignment and will be reported to the department head with further potential consequences. In addition, students are responsible for familiarizing themselves with the University's numerous policies and procedures contained in the University Catalog and Student Handbook. The Code of Conduct policies and the Policy on Academic Misconduct are of special significance, since cheating, plagiarism, and personal misconduct are strictly prohibited and carry severe penalties. Students should read and understand Eastern's Academic Misconduct Policy, which can be found in the student handbook.

www.easternct.edu/academicmisconduct

All violations will be handled under the procedures established in this policy.

Classroom civility

Cell phones are not appropriate in class and must be turned off or set to vibrate and stored off of the class desk. In general, follow the Golden Rule and treat others with respect and how you want to be treated.

Accommodations for Students with Disabilities

Eastern Connecticut State University is committed to following the requirements of the Americans with Disabilities Act (ADA) of 1990, the ADA Amendment Act of 2008, and Section 504 of the Rehabilitation Act of 1973, as amended in 1998. If you are a student with a disability (or think you might have a disability) and require accommodations or assistance evacuating a building in the case of an emergency, please contact the Office of AccessAbility Services (OAS) at 860-465-0189 to discuss your request further. Please note that accommodations are not retroactive and must be communicated through a Letter of Accommodation, which is drafted by the OAS.

*Tentative course schedule

Week	Week of	Topic
1	1/21/19	<i>MLK holiday - No Class Monday</i>
		Introduction to Big Data Programming, Docker, and Linux
2	1/28/19	Hadoop Architecture and HDFS
3	2/4/19	Basic Python Programming
4	2/11/19	The Map-Reduce Framework
		<i>Lincoln's Birthday - No Class Friday</i>
5	2/18/19	<i>President's Day - No Class Monday</i>
		Review / Exam I
6	2/25/19	Relational Databases and MySQL
7	3/4/19	Introduction to Sqoop, Impala, and HIVE
8	3/11/19	<i>Spring Break - No Class</i>
9	3/18/19	Data Formats and Data File Partitioning
10	3/25/19	Review / Exam II
11	4/1/19	Spark Basics, Functional Programming, and RDDs
		<i>Day of Reflection - No Class Friday</i>
12	4/8/19	Programming in PySpark
13	4/15/19	Aggregating Data with Paired RDDs
14	4/22/19	Spark RDD Persistence
15	4/29/19	Students work on Projects
16	5/6/19	Advanced Topics in Big Data Programming and Management
	5/17/19	Final Exam (Friday, 5/17, 11:00 – 1:00 PM)