CSC 369: Introduction to Distributed Computing
Winter 2019
Course Syllabus

January 6, 2019

Instructor: Alexander Dekhtyar
email: dekhtyar@csc.calpoly.edu
office: 14-210

Lecture	MWF	2:10 – 3:00pm	14-250
Lab	MWF	3:10 – 4:00pm	14-309

Office Hours

	When	Where
Monday	9:10am - 10:00am	14-210
Tuesday	1:10pm - 3:00pm	14-210
Wednesday	9:10am - 10:00am	14-210

Additional appointments can be scheduled by emailing the instructor at *dekhtyar@calpoly.edu*.

Overview

In this course we study the design and implementation of a variety of data processing algorithms on distributed computing frameworks.

Texbook

The course does not have a required textbook, primarily because no book known to the intructor has exactly the content covered in this course. However, there is an O'Reiley book for each component of our class. The books below are all recommended reading.

- Donald Miner, Adam Shook, *MapReduce Design Pattern: Building Effective Algorithms and Analytics for Hadoop and Other Systems*, O'Reiley Media, 1st Edition, 2012, ISBN: 978-1449327170.

- Mahmoud Parsian, *Data Algorithms: Recipes for Scaling Up With Hadoop and Spark*, O'Reiley Media, 2015, ISBN: 978-1491906187.
- Christina Chodorow, *MongoDB: The Definitive Guide*, O'Reiley Media, 2013, ISBN: 978-144924468
- Holden Karau, Andy Konwinski, Patrick Wendell, Matei Zaharia, *Learning Spark: Lightining-Fast Big Data Analysis*, Packt, 2015, ISBN: 978-1449358624
- Tomasz Drabas, Denny Lee, *Learning PySpark*, O'Reiley Media, 2017, ISBN-13: 978-1786463708

Topics

The following will be covered in the course.

No.	Topic	Duration (weeks)
1.	Introduction: Distributed Systems and Computations	1
2.	Key-Value Relationships / MongoDB	2
3.	MapReduce/ Hadoop	3-4
4.	Resilient Distributed Datasets and Spark	3-4

Most of the topics will be covered in the order specified above, but some variations are possible during the course.

Grading

Homeworks	0–5%
Labs	40 – 50%
Midterm Exams	45-50%

I give relatively hard problems and take points off on exams. Because of this, the traditional 90-A, 80-B, 70-C grading schema does not work in my classes. Historically, the A/B cutoff has been around 80-85%, while the B/C cutoff has been around 67-70%.

Course Policies

Exams

There will be two midterms and no final exam in this course (I may be away from the country during the finals week).

The midterms will probably take place on one of the following days: *February 11 (Monday)* or *February 13 (Wednesday)*, and *March 15, (Friday).*

Make-up exams will not be given, *unless there are extraordinary circumstances present and I am notified in advance.* The policy regarding the use of textbooks and notes will be announced at least one week prior to each exam.

Homeworks, Labs

The course will have 6-8 lab assignments, designed to let you test in practice what we have learned in class. Each lab assignment will span multiple lab sessions (typically 2 or 3). Due dates/times will be explicitly provided for each assignment. Typically, the assignments are due midnight of the day of the last lab period. Often a new lab assignment will already have been specified by then. You are welcome to work on the lab assignments outside the lab hours, however, lab period attendance is highly encouraged.

Some assignments will be individual. Other will be small team or pair programming assignments. Groups/pairings are to be formed by you - I will only intervene if someone cannot find a group/pair, or if there is a hard-to-resolve issue that requires my attention. You are welcome to stay in the same group/pair for multiple lab assignments, or form a new group/pair for each non-individual assignment. All members of a group will recieve the same grade for the assignment.

In addition to labs, a number of paper-and-pencil homeworks will be assigned. Homeworks are my way of providing you with some exam study guides. Homeworks will be collected, but will not, as a rule, be graded.

Late Submissions

All assignments are due at classtime on the due date: homeworks - at the beginning of the class (with grace period extending to the beginning of the lab period); lab assignments - at the end of the lab period. Any deviations from these rules will be spelled out explicitly in the assignments.

Homework/lab assignments submitted later than indicated above will be considered *late submissions*.

If paper-and-pencil homework solutions are distributed on the due date of the homework, ***late homework submissions will not be accepted***. Otherwise, late homeworks can be submitted during next 24 hours for a 10-30% penalty (the exact amount will depend on the submission time and the specific circumstances). No homework submissions will be accepted afterwards.

Late lab assignment submissions can be turned in before or at the beginning of the next lab period for a 10-30% penalty (the exact amount will depend on the submission time and the specific circumstances[1]). No lab assignment submissions will be accepted after that.

Communication

The class has an official mailing list:

csc-369-01-2192@calpoly.edu

[1] The penalty will be larger if the gap between the two lab periods includes a weekend and smaller otherwise

All students enrolled in the class are automatically subscribed to the mailing list.

I encourage questions during classtime and questions via email. My answers to email questions may be broadcast to the entire class via the mailing list, if the answer may be relevant to everyone (e.g. a correction in a text of a handout, or a clarification of a homework problem), and may also appear on the web page. The questions can also be posted to the mailing list directly. The mailing list will also be used for all annoucements related to the course. It is your responsibility to read your class-related email. Failure to read email posted to mailing list cannot be used as an excuse in the class.

Web Page

Class web page can be found at

http://www.csc.calpoly.edu/∼dekhtyar/369-Winter2019

Through this page you will be able to access all class handouts including homeworks, project information and lecture notes (should the latter be written).

Academic Integrity

University Policies

Cal Poly's Academic Integrity policies are found at

http://www.academicprograms.calpoly.edu/academicpolicies/Cheating.htm

In particular, these policies define *cheating* as (684.1)

> "*...obtaining or attempting to obtain, or aiding another to obtain credit for work, or any improvement in evaluation of performance, by any dishonest or deceptive means. Cheating includes, but is not limited to: lying; copying from another's test or examination; discussion of answers or questions on an examination or test, unless such discussion is specifically authorized by the instructor; taking or receiving copies of an exam without the permission of the instructor; using or displaying notes, "cheat sheets," or other information devices inappropriate to the prescribed test conditions; allowing someone other than the officially enrolled student to represent same.*"

Plagiarism, per University policies is defined as (684.3)

> "... the act of using the ideas or work of another person or persons as if they were one's own without giving proper credit to the source. Such an act is not plagiarism if it is ascertained that the ideas were arrived through independent reasoning or logic or where the thought or idea is common knowledge. Acknowledgement of an original author or source must be made through appropriate references; i.e., quotation marks, footnotes, or commentary."

University policies state (684.2): "Cheating requires an "F" course grade and further attendance in the course is prohibited." (appeal process is also outlined, see the web site above for details.). Plagiarism, per university policies (684.4) can be treated as a form of cheating, although a level of discretion is given to the instructor, allowing the instuctor to determine the causes of plagiarism and effect other means of remedy. It is the obligation of the instructor to inform the student that a penalty is being assessed in such cases.

Course Policies

All homeworks are to be completed by each student **individually**. Lab assignments are to be completed by the appropriate units (individual, pair, group), and no code/solution-sharing between units is permitted. Students are encouraged to discuss class content among themselves but NOT in a manner that constitutes plagiarism and cheating as defined above (e.g., you can solve together a problem from the textbook that had not been assigned in the homework, but you should solve assigned problems individually).