Data Intensive Computing (CS 491/691-1006)

Spring 2018

# Course Information

## Instructor Information:

**Instructor:** Dr. Engin Arslan  
**Office:** SEM 238  
**Phone:** (775) 784-6917  
**Email:** earslan@unr.edu   
**Office Hours:** Monday 2—4pm

**Time and Room:** TuTh 3:00PM - 4:15PM at PE 105

## Course Description:

Data-intensive computing has been receiving much attention as a collective solution to address the data deluge that has been brought about by tremendous advances in distributed systems and Internet-based computing. An innovative programming models such as MapReduce and a peta-scale distributed file system to support it have revolutionized and fundamentally changed approaches to large scale data storage and processing. These data-intensive computing approaches are expected to have profound impact on any application domain that deals with large scale data, from healthcare delivery to military intelligence. Given the omnipresent nature of large scale data and the tremendous impact they have on a wide variety of application domains, it is imperative to ready our workforce to face the challenges in this area. This course provides a broad introduction to the fundamentals in data intensive computing and its enabling systems architectures such as MapReduce, Spark, Hbase, and Hive.

## Course Pre/Co-requisites:

Introduction to programming or data structures/ algorithms, computer architecture, or instructor approval.

## Required texts, course materials:

There is no required books but following two books are suggested:

* Jimmy Lin and Chris Dyer, [Data-Intensive Text Processing with MapReduce](http://www.umiacs.umd.edu/%7Ejimmylin/book.html). 2010. (see also the [final pre-production manuscript](http://www.umiacs.umd.edu/%7Ejimmylin/MapReduce-book-final.pdf))
* Tom White, [Hadoop: The Definitive Guide](http://www.hadoopbook.com/), O'Reilly, 2010. (**Second Edition**) [[pdf](http://fivedots.coe.psu.ac.th/%7Esomchai/Download/Cloud%20Model/Knowledges/E-Book/Hadoop.pdf)]

## Student Learning Outcomes:

This course will help students achieve the following learning outcomes.

* An ability to apply knowledge of computing, mathematics, science, and engineering.
* An ability to analyze a problem, and identify, formulate and use the appropriate computing and engineering requirements for obtaining its solution.
* An ability to use current techniques, skills, and tools necessary for computing and engineering practice.

## Course Requirements:

There is a semester-long course project done in four parts. Details will be presented in class.

This is a tentative grading division.

* Project & Presentations: ~ 50%
* Homework assignments: ~ 25%
* Midterm exam : ~ 25%

## Grading Criteria, Scale, and Standards:

Enter grading policy, including statement on whether or not plus/minus grading will be used, and list letter grade assignment.

Sample letter grade assignment:

A: 93% - 100%

A-: 90% - 92.9%

B+: 87% - 89.9%

B: 84% - 86.9%

B-: 80% - 83.9%

C+: 77% - 79.9%

C: 74% - 76.9%

C-: 70% - 73.9%

D+: 67% - 69.9%

D: 64% - 66.9%

D-: 60% - 63.9%

F: <60%

\* Late and missed assignments/exams may be accepted (with cause) at the discretion of the instructor.

## Course Topics (Tentative)

* What is data intensive distributed computing?
* Distributed Filesystems and NoSQL databases, e.g., HDFS and HBase.
* Resource management, e.g., YARN and Mesos.
* Execution engines, e.g., MapReduce and Spark.
* Query/scripting languages, e.g., Hive.
* Graph processing, e.g., Pregel, GraphLab, PowerGraph and GraphX.
* Streaming processing, e.g., Spark Streaming.

# University Policies

## Statement on Academic Dishonesty:

"Cheating, plagiarism or otherwise obtaining grades under false pretenses constitute academic dishonesty according to the code of this university. Academic dishonesty will not be tolerated and penalties can include filing a final grade of "F"; reducing the student's final course grade one or two full grade points; awarding a failing mark on the coursework in question; or requiring the student to retake or resubmit the coursework. For more details, see the [University of Nevada, Reno General Catalog](http://catalog.unr.edu/)."

Statement of Disability Services:

Any student with a disability needing academic adjustments or accommodations is requested to speak with me or the [Disability Resource Center](http://www.unr.edu/drc) (Pennington Achievement Center Suite 230) as soon as possible to arrange for appropriate accommodations.

**This course may leverage 3rd party web/multimedia content, if you experience any issues accessing this content, please notify your instructor.**

## Statement on Audio and Video Recording:

"Surreptitious or covert video-taping of class or unauthorized audio recording of class is prohibited by law and by Board of Regents policy. This class may be videotaped or audio recorded only with the written permission of the instructor. In order to accommodate students with disabilities, some students may have been given permission to record class lectures and discussions. Therefore, students should understand that their comments during class may be recorded."

**The University of Nevada, Reno is committed to providing a safe learning and work environment for all. If you believe you have experienced discrimination, sexual harassment, sexual assault, domestic/dating violence, or stalking, whether on or off campus, or need information related to immigration concerns, please contact the University's Equal Opportunity & Title IX office at 775-784-1547. Resources and interim measures are available to assist you. For more information, please visit the** [**Equal Opportunity and Title IX**](https://www.unr.edu/equal-opportunity-title-ix) **page.**