Course:	Big Data Analysis		
Course type:	Elective	EE, CE, and IT*	Credit: 3
Level:	Undergraduate- Graduate		
Co-requisite(s):			
Prerequisite(s):	At least one of the following courses: Database, Data Mining, Machine Learning, Introduction to Big Data		
Prerequisite by topic:	Familiarity with Java and Python.		
Textbook(s):			
Coordinator:			
Goals:	The recent explosion of IT services has resulted in creation of large volumes of mostly unstructured data: service logs, sensor data, images, voice data and similar. Today, we have the hardware to reliably and cheaply store huge volumes of data. In order to efficiently analyze data and extract useful information, a general knowledge of big data and its tools and techniques is needed. This course introduces students to key concepts in big data analytics including: data scalability, data science, cloud computing and super computing. The students will also explore a few applications of big data that arise in practice including data centers, real-time services and deep learning. This course reviews technologies that let students manipulate, store, and analyze big data. We will look at the basic analysis tools including MATLAB, R and Python. We will review MapReduce techniques for processing large quantities of data and Hadoop, a framework that effectively implements MapReduce on Internet scale problems. We will study highly inter-connected super-computers and will explore the advantages and disadvantages of super-computing in contrast with could		

	computing.
Outcome:	Participant is expected to have a broad knowledge of tools to process big data. Participants are expected to become familiar with: - MATLAB, Python, R - Multi-threading and parallel processing - GPU programming - MapReduce and Hadoop - Elastic Search, Logstash and Kibana - Storm - Message Passing Interface - Super-Computers - Parallel-algorithms - Machine Learning and Deep Learning - Real-time Internet Services
Topics:	**Introduction** - Why/when big data is needed? - Challenges of Big Data **Single Computer** - Data Processing with Matlab, Python and R - Multi-threading and OpenMP - GPU and CUDA **Cloud** - MapReduce framework and Hadoop - Cluster Computing with Spark - Real-time analytics with Storm **Super Computer** - Message Passing Interface - High Performance Computing - Large Scale Storage and Backup Systems **Applications** - Data Centers - Distributed Search - Large Scale Machine Learning - Real-time Services - Future Trends
Computer usage:	Students need access to a computer with a 64-bit operating system and at least 4 GB of RAM. 8 GB of RAM is recommended. Students need to setup and use various big data tools. Probably the students will be able to use a commercial cloud

	service.
Assignments:	Students will have practical/programming assignments to experience a number of big data technologies.
Projects:	The students need to implement a big data platform. This platform must have at least one of the following characteristics: 1- Be real-time, 2- Perform machine learning, 3- Be distributed, 4- Provide a web-service
Grading:	Assignments 20% Quizes 30% Final Exam 20% Final Project 30%
Further readings:	
Prepared by:	Mohammad Amin Sadeghi
Date:	OCT 25, 2016