CS590RA Real Time Big Data Analytics

Credits: 3
CSEE - UMKC

Course Coordinator: Yugyung Lee

Graduate Courses:

Frequency of offering: Every Fall

Specifics for Fall 2014:

Class Times and Room: Class Hours:
T/TH 1:00 – 2:15PM in FH462

Instructor Identifiers:
Instructor: Yugyung (Yugi) Lee
Office Hours: T/Th 4:00 – 5:00PM or by appointment
Office: FH 560D
Phone: 816-235-5932
Email: leeyu@umkc.edu
Website: www.csee.umkc.edu/~leeyu

TA: Feichen Shen
Office: FH521
Email: fsm89@mail.umkc.edu
Office Hour: TBA

Course description:
This course teaches students fundamental theory and practice in the field of big data analytics and real time distributed systems for real time big data applications. Especially, this course provides students with hands-on experience for the development of real-time applications with real time stream data such as twitter messages, Instagram images.
Prerequisites: CS451. Recommended preparation: knowledge or programming experience with cloud computing and big data analytics.

No Textbook, but other Materials Required:
o   Research papers (Class Handout)
Books Recommended
Real-Time Big Data Analytics: Emerging Architecture
By Mike Barlow, Publisher: O'Reilly, Released: June 2013
Mining of Massive Datasets
By Anand Rajaraman and Jeffrey David Ullman
Publication Date: December 30, 2011 | ISBN-10: 1107015359 | ISBN-13: 978-1107015357
Data-Intensive Text Processing with MapReduce
By Jimmy Lin and Chris Dyer

Course Objectives, Goals or Learning Outcomes:
Stream processing is ingest real-time big data per second and analyze it in near real time. This course explores technologies and tools that can be used for real-time big data analytics and also provides students with hands-on experience with the real-time application development for big data analytics. Throughout this course, students will obtain extensive hands-on experiences in the development of real-time big data application using big data analytics techniques and with various tools such as Twitter's Storm, Apache Flume, Apache Kafka, Google's BigQuery. Students will be able to apply the concepts and techniques to emerging applications such as real-time mobile applications with big data.

Instructional Strategies/Pedagogical Approach:
This course will require several distinct types of learning: Since this course is a research-oriented graduate course, a substantial portion of the quarter will be devoted to student presentations of techniques and research papers in the areas of Big Data Analytics. Students will be expected to select a problem area in Big Data Analytics and prepare an intensive presentation covering the methods and framework commonly employed to address their problem.

a. Research Project: Students will be asked to design and build an innovative research project for presentation at the end of the semester. Students should organize themselves into research project teams and develop their research project. A final written report will be submitted.

b. Reading/Discussion/Presentation: The lecture/discussions are designed to be highly participatory. Therefore, it is fair and just that points are awarded for effort and participation in these discussions. For each research paper in the assigned reading list: participate in the class discussion of each paper provide written summaries of each paper before class volunteer to present in class certain of the papers on the reading list, on a rotating basis.

Semester Schedule of Topics ("subject to modification"):
1. Introduction to Real-time Big Data Analytics
2. Big Data Platform/Systems
   - Cloudera Platform
   - MapReduce/HDFS
   - HBase/Hue
3. Real Time Big Data Platform/Systems
   - Twitter Storm
   - Kafka/Stark/Flume/ZooKeeper
   - Google BigQuery
4. Data Processing and Analytics with Machine Learning
   - Machine Learning for Big Data
   - Machine Learning for Text data (Twitter messages)
   - Machine Learning for Image data processing (Instagram images)
5. Intelligent Real-Time Applications
   - Large Scale Mobile Applications
   - Real Time Mobile Apps
   - Activity Recognition and Remote Sensing

Research Projects
Students will be asked to build/create an innovative research project for presentation at the end of the semester. Students will form teams of 2 - 3 members and work on projects as a team within a particular track. Teams and projects will be decided according to the timeline below. Students are welcome to formulate their own project ideas. Each team will be required to present their project to the class and a final project report written in the style of a conference paper will be handed in following the presentation.

Important Dates for the Semester:
o Sep. 1 (M): Labor Day Holiday - University Closed
o Nov. 24 (M) – Nov. 28 (F): Thanksgiving Holiday
o Dec. 12 (F): Last Day of Class

Timelines for Project Development: (Tentative)
o Proposal/Plan, Incremental Project Reports: TBA
o Project Presentations: Dec. 9 (T) & Dec. 11 (Th)
o Final Project Report: Dec. 12 (F)

Technical Challenges
This course aims to provide a hands-on learning experience so that students apply their practical knowledge and skills for the development of real-life, data-intensive applications. Two or three technical challenges are designed for students to compete with each other in terms of developing of the best solution for given challenges. Real challenges in fields such as analysis of social media, business, and biomedical data are planned. The students will have practical learning and hands-on experiences through the use of Real-time Real World Data, Machine Learning algorithms and infrastructure used by Real Time Big Data technologies.

Tutorials will be given:
1. Tutorial 1: Aug. 28 (Th): GitHub, Agile Tool, VMware
2. Tutorial 2: Sep. 2 (T): Android App/Rest Web Services
3. Tutorial 3: Sep. 4 (Th): Cloudera/MapReduce
4. Tutorial 4: Sep. 9 (T): Cloudera/HBase/Hue
5. Tutorial 5: Sep. 11 (Th): Twitter Storm/ZooKeeper
6. Tutorial 6: Sep. 16 (T): Twitter Storm/Flume/Kafka
7. Tutorial 7: Sep. 18 (Th): Twitter Storm/Flume/BigQuery
8. Tutorial 8: Sep. 23 (T): Machine Learning with R/Mahout (1)
9. Tutorial 9: Sep. 25 (Th): Machine Learning with R/Mahout (2)
10. Tutorial 10: Nov. 20 (Th): Project Video Production

Paper Reviews and Presentations
Students are required to read, present, and discuss graduate-level research papers throughout the semester. A paper per week will be read and reviewed. Written reviews of each paper to be discussed in class are due prior to the start of that class, and should be post to students' website. Late reviews will not be accepted. Each paper to be discussed in class will be assigned to a student to present in class. Assignments will rotate throughout the class. Papers will be assigned approximately one week in advance of the presentation date. The presenter of a given paper must submit their Power point slides to the blackboard system by midnight of the night before the presentation.

Assignments

Students may need to complete reading assignments, exercises, problem sets, review case studies, and engage in implementation and research tasks on Knowledge discovery and management. The late policy on assignments is 10% off the grade if late within one day, 20% off the grade for two days late, 30% off the grade for three days late. Assignments that are submitted more than three days late will no longer be accepted. More information will be available on the Announcements web page.

Evaluation / Assessment Criteria and Grading:

Project (Proposal/Reports/Presentation)	30%
Journal Club (Critiques, Presentations, Discussion)	20%
Technical Challenges	20%
Exam	20%
In-Class Activities & Attendance	10%

Grading policy: 100 - 94 = A; 93 - 90 = A-; 89 - 87 = B$^+$; 86-84 = B; 83 - 80 = B$^-$ and so on.

Late Submission:

The late policy on assignments is 10% off the grade if late within one day, 20% off the grade for two days late, 30% off the grade for three days late. Assignments that are submitted more than three days late will no longer be accepted. More information will be available on the Announcements web page.

Policy on Student Attendance and Make-ups:

Each student should make every attempt to get to class on time. The instructor is willing to circulate a sign-in sheet at every class and missing more than two class sessions may result in a reduced grade. With the exception of documented emergencies, medical reasons or out of town travel related to work, make-ups will not be possible. Whenever possible, advance notification is required.

University Policy on Student Conduct:

Cheating, plagiarism, disruptive behavior and other forms of unacceptable conduct are subject to strong sanctions in accordance with university policy. See detailed description of university policy at the following URL: http://www.umkc.edu/html/handbook/policies-and-regulations/conduct.html

University policy on English proficiency of Instructors:

"Students who encounter difficulty in their courses because of the English proficiency of their instructors should speak directly with their instructors. If additional assistance is needed, they may contact the UMKC Help Line at 816-235-2222 for assistance."