Syllabus of ITCS6010: Big Data Cloud Deployment

Spring 2017, CS Dept / College of CCI

Time & Loc.: Mon. 6:30pm -- 9:15pm, Woodward Hall 135

Instructor:

Prof. Dazhao Cheng Office: Woodward Hall 435G, (704) 687-8381 (office), Email: dazhao.cheng@uncc.edu

Office Hours: T 2:30 pm-4:30pm, Woodward Hall 435G, and/or by appointment (email preferred)

Teaching Assistant: Ashish Raju Mahendrakar
Office hours: Thursday 2:00 pm - 5:00 pm; Email: amahendr@uncc.edu

Course Description:

ITCS6010 is a tour through latest techniques in various big data systems on clouds, covering topics in cloud computing, autonomic computing and sustainable computing. We will explore solutions and learn design principles for building large distributed computational systems (e.g., Apache Hadoop and Spark) to support data-intensive computing on clouds.

Building large scale distributed systems that support data-intensive computing involves challenges at multiple levels, from the network (e.g., transport, routing) to the algorithmic (e.g., data distribution, resource management) and even the social (e.g., incentives). This course encompasses a comprehensive study of the system architecture, enabling technologies, software environment, and innovative applications of data-intensive computing on clouds. The course will give students an overview of the inner workings of the open source cloud system, server virtualization techniques, big data processing frameworks (e.g., Apache Hadoop, Spark and Spark Streaming), the MapReduce programming model and their applications. In addition, we will discuss recent research in cloud computing and big data with a focus on efficient cloud resource management, performance control, and energy efficiency. The course will require student-led paper presentations and discussions. Students will work on cloud computing and data-intensive application related projects in groups and write a report in a research conference format. The course covers fundamental concepts and principles of big data clouds with exercises which guide you to apply the big data theory and design principles, verify their understandings, and build a solid foundation for creating innovations in today's data-intensive applications.

Course Format

The material presented in the course will include the following textbooks.

Suggested Textbooks:

1, Hadoop: The Definitive Guide, 4th Edition, By Tom White, Publisher: O'Reilly Media
Final Release Date: March 2015
2, Learning Spark Lightning-Fast Big Data Analysis, By Holden Karau, Andy Konwinski, Patrick Wendell, Matei Zaharia, Publisher: O'Reilly Media, Final Release Date: January 2015

Tentative Schedules

- Cloud computing: current & future trends
- Server virtualization of compute, storage and network
- Cloud operating system, nova scheduler, object storage
- Hadoop architecture, HDFS and Y ARN
- MapReduce programming model and scheduling algorithms
- Spark for lightning fast cluster computing
- Spark Streaming for real-time streaming data processing
- Big Data security and privacy challenges in the cloud
- Performance benchmarks, analysis & tuning

Prerequisites

Programming and Unix/Linux environments (or equivalent). ITCS3143/ITCS3146 (Operating System) or equivalent. Knowledge of UNIX/Linux systems and programming skills (C, C++, or Java). If you want to take the class without the prerequisite, you need to get the approval from the instructor and make up for the prerequisite.

Grading

The final grade will be composed of

- In-class attendance 5%
- In-class discussion 10%
- Project 1 (Survey Report & Reading Assignments) 15%
- Homework Assignments (Technique Report) 20%
- Presentation (in class by individual) 20%
- Final Project 2 (Technique Report, comprehensive) 30%

Grades will be assigned as follows:

- 90≤{A}; 87≤{A-}<90
- 84≤{B+}<87; 80≤{B}<84; 77≤{B-}<80
- 74≤{C+}<77; 70≤{C}<74;
- 65≤{D+}<70; 60≤{D}<65; E/F:below60

Requirements

1. Students are required to attend all lectures.
2. The University policy on Course Withdrawal allows students a limited number of opportunities available to withdraw from courses. There are financial and academic consequences that may result from course withdrawal.
3. Homework/reading/project assignments are important part of the course and are to be completed individually. There will be about three homework assignments, one reading assignment, one presentation assignment and one team project. Homework must be done individually, and be due in class on the due date. The reading assignment and the project should be done in two-person teams; your teammates will possibly be designated by the instructor (based on random selection) before projects are released. If a teammate is not available, a project might be done individually, but should be approved by the instructor in advance. Demos and reports for the project are required.
4. Late homework/reading/project submissions lose 30% of their values per day, except under extreme non-academic circumstances, such as illness. In such cases, you have to inform the instructor by email/phone right away and provide sufficient and convincing proof later, i.e., documents from the doctors.
5. FOR FAIRNESS, NO MAKE-UP EXAMS, exceptions are the same as those of late homework.
6. There will be one midterm exam and one final exam, which are open-book and open-notes.

Others

If you have a disability for which you are requesting an accommodation, you are encouraged to contact the Disability Services Office. Cheating, unfortunately, it is necessary to mention it here. Cooperation is not the same as cheating. It's OK to ask someone about the concepts before you start to do homework or project assignments; however, copying other people's code or solution sets is strictly prohibited. Any work submitted for a grade must include the following statement and be signed and dated. If this is missing or not signed and dated, the work will be returned un-graded.

We need the strict rules, because everyone wants to be, and will be, treated fairly in this class!