

Special Topic: EEL 4932: Massive Storage and I/O for Big Data Computing (Spring 2015)

Syllabus

Instructor: Dr. Jun Wang, HEC 320, 823-0449(office), jun.wang@ucf.edu.
Classroom and Time: ENGRI-384 on MW 13:30-14:45.

Web: <http://www.eecs.ucf.edu/~jwang/>

Course Objective and Description:

This course provides a broad introduction to the fundamentals of massive file storage systems and I/O architectures in big data computing and its enabling systems Infrastructure such as MapReduce and storage, with a focus on system architecture, file storage, distributed file systems and parallel file systems, programming models, application development and evaluation. Selected scientific applications will be used as case studies.

Prerequisite: EEL3801/CDA3103 computer organization, or EEL4768 computer architecture, or instructor approval.

Required textbooks:

- Computer Architecture: A quantitative approach, 5th edition, Hennessy&Patterson, Morgan Kaufmann, September 2011

Reference textbooks:

- Hadoop: The Definitive Guide (3rd Edition), Tom White, O'Reilly Media, 2012.
- Many recent papers in leading conferences/journals will be discussed.

Course Outline (tentative):

1. Introduction to Big Data Computing, and its enabling systems architecture such as Data Center and Cloud
2. Performance, Efficiency and Energy Evaluation and Analysis Methods
3. Introduction to Storage System
4. RAID
5. File System and Database

6. Distributed and Parallel File Systems
7. Warehouse Scale Computers
8. Data Parallel Programming Models
9. Introduction to Hadoop
10. MapReduce Runtime Management
11. Case Study

Grading Policies:

- Final Grade will be calculated according to the weight associated with each component listed:
 - Homework Assignments: 25%;
 - Class Participation: 10%;
 - Class Presentation: 25%;
 - Class Project: 15%;
 - Final Exam: 25%;
 - **Grading Scale: $A \geq 90$; $80 \leq B < 90$; $70 \leq C < 80$; $60 \leq D < 70$; $F < 60$**

Note: Homework and programming assignments are due by 11:59pm of the due date (unless announced in class otherwise). Late homework (non-programming) will NOT be accepted. Late program penalty is 10% per day, according to the timestamp of your online submission. Only when verifiable extenuating circumstances can be demonstrated will extended assignment due dates be considered. Verifiable extenuating circumstances must be reasons beyond control of the students, such as illness or accidental injury. Poor performance in class is not an extenuating circumstance. Inform your instructor of the verifiable extenuating circumstances in advance or as soon as possible. In such situations, the date and nature of the extended due dates for the assignments will be decided by the instructor.

Attendance Policy:

Attendance is required. Students are responsible for any material covered in class. Lots of the materials covered in class will not be in the textbook. Announcements about homework, projects, programming assignments, etc. may be made in class or online or by emails. Students are encouraged to check the online WebCourses regularly.