CS 292 Special Topics on Big Data

Instructor: Yuan Xue (yuan.xue@vanderbilt.edu)

Course description:
This course provides an introduction to the principles and practices of large-scale data processing and analysis. Topics include: data storage (e.g., HDFS), SQL vs. NoSQL (e.g., HBase) database, data modeling, data processing and query (e.g., MapReduce/Hadoop, Dremel/Impala/Drill, Storm, Spark), data analysis (e.g., Mahout) and high-level workflow management tools.

This course is for both senior undergraduates and graduate students. It focuses on case studies and programming assignments, where students could hand on some of the powerful open source tools for big data management and analysis. The course requires students to conduct a semester-long project. Individualized advices are given for project topic selection, design, report preparation and presentation.

Credit: 3 credit hours
[Graduate students can take it for graduate credits with extra work]

Prerequisite
The course requires basic programming skills (Java preferred) and knowledge on Linux system. Background on database, computer network and operating system is preferred but not required.

© 2013 Yuan Xue