

COURSE SYLLABUS

Big Data Computing

Instructor

Name, title: Zoltan Toth, Visiting Professor
Department: Department of Economics / CEU Business School
Semester/term, year: Winter AY 2015/2016
Course level: MSc
Credits: 2 credits (4 ECTS)
Program: MSc in Business Analytics

Course Description

This course examines the basic concepts and practices of big data computing. We will cover the challenges that arise when the size of data to be analyzed outgrows the limits of traditional data-analytics systems, the new challenges big data computing introduces and the evolution of the big-data ecosystem. We will touch upon classical subjects such as MapReduce and modern approaches as Spark and big data computing in the cloud.

Tentative outline

1. Introduction, the challenges of distributed data analytics
2. The Big Data Ecosystem, Hadoop, MapReduce
3. YARN, HDFS deep dive, MapReduce design patterns
4. The Hadoop Streaming API
5. Pig
6. Hive
7. Spark
8. Spark ML - Machine Learning on Big Data
9. Big Data Computing in the cloud: Amazon Web Services, Elastic Mapreduce
10. Real-time data processing / Streaming solutions
11. Data Warehousing in the cloud: Amazon Redshift

Assessment

Assessment is through a final exam (40%), and assignments (60%). Assignments will be done in teams. You will have to implement big data analytics applications in the cloud.