Syllabus

Department of Economics and Business
CENTRAL EUROPEAN UNIVERSITY

Data Infrastructure in Production

- **Instructor:** Zoltán Tóth, Gergely Daróczi, Ryan McCabe (TA)
    - office hours: right after classes and by appointment on skype (zoltanctoth)
- **Credits:** 2 (4 ECTS)
- **Term:** Spring 2017-2018
- **Course level:** MSc
- **Prerequisites:** Data Analysis 1a; Big Data Computing

Course availability

The course is core for MS in Business Analytics students. The course may not be taken for other students due to a cap required by IT need.

Course description

In this course, you will learn what components make up a data architecture in production and which technologies are required for different use-cases. We will cover different data processing and database technologies and see how they can be used for solving a variety of problems. Will take a look at these using cloud computing using AWS (Amazon Web Services). You will also learn about best practices on how to tackle real-world business analytics problems using these technologies and how to deploy business rules implemented in R into production as part of cloud-based stream-processing engines, dashboards or scoring APIs.

Learning outcomes

By the end of the course you will:

- Understand the building blocks of a production data infrastructure.
- You will have an overview of current data-related cloud computing services.
- You will have hands-on knowledge on how to build a simple, end-to-end data pipeline on Amazon Web Services (AWS).
- You will have hands-on knowledge on creating dashboards in R on the top of AWS.

Reading list

Data, presentations and code for the exercises will be provided.

There is no compulsory reading for the class. Suggested reading:

- Jay Kreps: I heart logs
- Martin Kleppmann: Designing Data-Intensive Applications

Assessment

- 20% quizzes at the beginning of class
- 80% final project – you will need to create a data pipeline on AWS

Grading policy

- Students shall not miss more than 1 day of lectures (out of 4 days). Failing to do so will yield an administrative fail grade. (If you have a major impediment please contact the Instructor.)
- To pass, students will need to get at least 50% of the overall grade. Failure to do so, will yield a Fail grade.

Course schedule and materials for each session

1. Overview of the database technologies on the top of Amazon Web Services

2. How to set up orchestration and data-management in a data infrastructure

3. Real-time data Processing with R and Amazon Kinesis

4. Creating interactive Dashboards with R and Shiny