Syllabus

Data Engineering 4: Using R in Production

- **Instructor:** Gergely Daroczi
- **Credits:** 1 (2 ECTS)
- **Term:** Winter 2019-2020
- **Course level:** MSc
- **Prerequisites:** Data Analysis 2: Finding Patterns with Regressions – Business Analytics track, Data Visualization 2: Practical Data Visualization with R; Data Engineering 3: Real-Time Data Processing
- **Course drop:** Course can be dropped free of charge 24 hours after the first session. After this date drop is possible until the course is halfway over (late drop fee applies). No changes are allowed past that date.

Course description

In this course, you will learn what components usually make up an automated, maintainable and scalable R-based data architecture, how to use that in production, and what other technologies you should become familiar with for different use-cases. The class starts with deploying real-world examples of business rules implemented in R onto cloud-computing resources running in AWS (Amazon Web Services), then will look into stream-processing with R, creating dashboard an scoring APIs. We will also cover some related data processing and database technologies outside of R: how they can be used for solving a variety of problems and how to connect to these from R.

Learning outcomes

By the end of the course you will have a better understanding on the building blocks of an R-based production data infrastructure in the cloud, including automated reports, dashboards, stream-processing and service integrations.

Reading list

Class materials will be available on GitHub.

Assessment

20% quizzes (at the beginning of the classes) and 80% final project (creating an R-ased data pipeline on AWS).

Grading Policy

Students shall not miss more than 1 day of classes, failing to do so will yield an administrative fail grade. To pass, students will need to get at least 50% of the quizzes AND at least 50% of the final exam.