Distributed Data Infrastructures, Fall 2018, Introduction

Jussi Kangasharju

Course topic and goals

- Distributed data infrastructures
- Design and hands-on
- After the course, you:
  - Knows different infrastructures and systems for large-scale data science processing
  - Can compare various infrastructures and their suitability for a particular problem
  - Can select the appropriate tools and environments for a particular problem
  - Can justify the system design choices behind existing data science infrastructures
  - Is able to implement or extend components for processing infrastructures

Course Organization

- Meetings: Tue 12-14, Thu 12-14 both in Exactum C222
- Article discussions and exercise Q&A

- Lecturer: Jussi Kangasharju
- Assistants: Nitinder Mohan, Otto Waltari, Aleksandr Zavodovski

Passing the Course

- No course exam, only exercises
- **10 essays based on scientific articles (max 2 points each)**
    - 1 point for summary of article
    - 1 point for exceptional additional insights
- Participation in class discussions (1 point per session)
    - Only if you returned corresponding essay
- **2 projects on different infrastructures (9 points each)**
- Half of maximum points for passing
- All assignments **in bold** are mandatory to return
- All assignment deadlines are strict, no extensions will be given

Course Timeline

- Week 1: Lecture on Tue and Thu
- Week 2: Return essays 1 (Mon) and 2 (Wed), start of project 1 (Tue)
- Week 3: Return essays 3 (Mon) and 4 (Wed)
- Week 4: Return essays 5 (Mon) and 6 (Wed), project 1 deadline (Sun)
- Week 5: Return essays 7 (Mon) and 8 (Wed), start of project 2 (Tue)
- Week 6: Return essay 9 (Mon), holiday on 6.12.
- Week 7: Return essay 10 (Mon), Q&A (Thu) project 3 deadline (Sun)

Practical Details

- See course page for info

- Announcements on Twitter #UnivHelsinkiCS_DDI18

- Slack team: See sign-up info on course page

- All returns via Moodle, see link on course page

- Office hours/appointments: Via email or ask in person after lectures

Questions?