Distributed Data Infrastructures, Fall 2017, 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 Physicum E204
• Tuesday sessions: Discussions, new material
• Thursday sessions: Q&A for exercises
• Lecturer: Jussi Kangasharju
• Assistant: Nitinder Mohan
Passing the Course

• No course exam, only exercises
• 5 essays based on scientific articles (1 point each)
• Participation in class discussions (1 point per session)
• 3 projects on different infrastructures (10 points each)
• Summary report/study diary (5 points)
• Half of maximum points for passing, 5/6 gets a 5
• 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, intro to AWS (Thu)
- Week 2: Return essay 1 (Mon), start of project 1 (Tue)
- Week 3: Project 1 deadline (Sun)
- Week 4: Return essay 2 (Mon), start of project 2 (Tue)
- Week 5: Return essay 3 (Mon), project 2 deadline (Sun)
- Week 6: Return essay 4 (Mon), start of project 3 (Tue)
- Week 7: Return essay 5 (Mon), project 3 deadline (Wed), summary report deadline (Sun)
Practical Details

• See course page for info
• Announcements on Twitter #UnivHelsinkiCS_DDI17
• 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
Setting up an AWS account

• We will use AWS for some of the projects
• Get an AWS Educate account:
  • https://www.awseducate.com/registration
• Install the CLI:
  • http://docs.aws.amazon.com/cli/latest/userguide/installing.html
• Try these before Thursday this week
• There will be more help with these on Thursday
Questions?