

Principles of E-Commerce I - Big Data Platforms

# Assignments 1+2

Todor Ivanov

Dr. Karsten Tolle

<http://www.bigdata.uni-frankfurt.de/>



## Assignments 1+2

- Assignments are done in groups of 2 persons (one exception of 3 persons possible)
- ... same group for the first two assignments
- Results (code & description) need to be submitted latest by **May 24, 2015 (23:55h)** – one submission per group!
- Submission of Assignment 1+2 together as one ZIP file!
- Results need to be presented on: **May 26 and 27, 2015** (about 15 minutes presentation time)

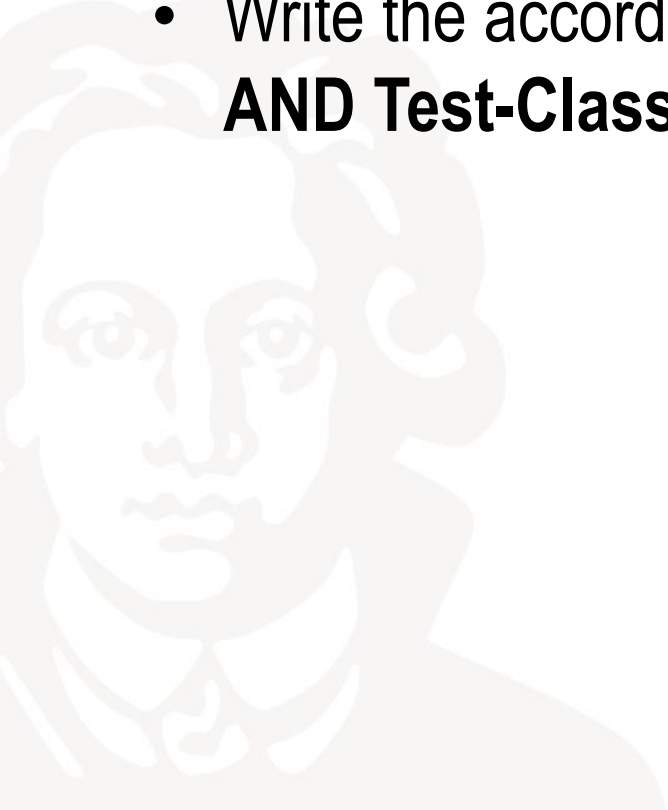
## Submissions via OLAT

- Registering: <https://olat.server.uni-frankfurt.de/olat/url/RepositoryEntry/3031171083/CourseNode/91531840833533>
- ... only one person per group!



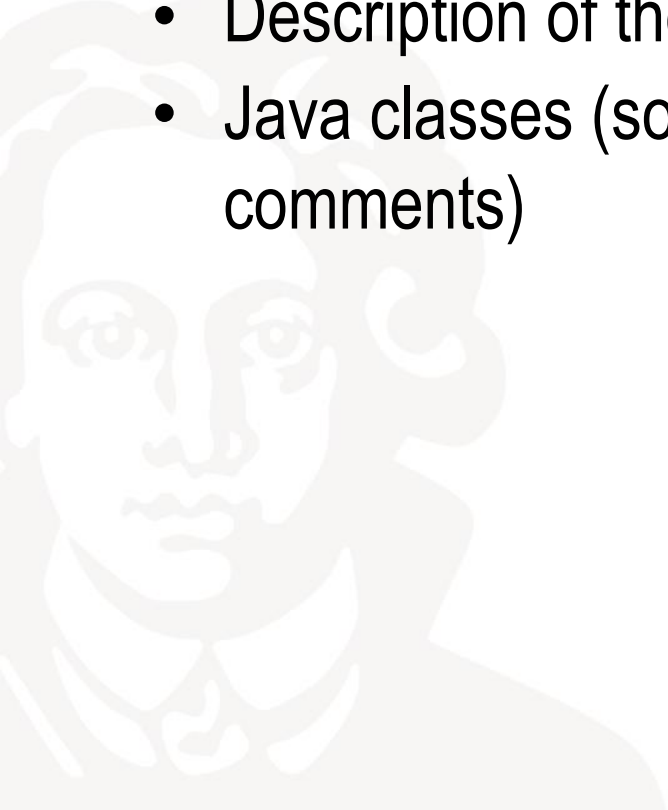
## Assignment 1 – Your own MR-Job

- Define your own MR-Job on existing data (e.g. from <https://www.govdata.de/> or <https://offenedaten.de/> ) that is meaningful (and goes beyond WC!).
- Write the according Java classes: Driver, Mapper, Reducer **AND Test-Class!!** (all with appropriate comments)



## Assignment 1 – Submission

- Data of Group members including: „**Name, Matrikelnummer und Studiengang**” (needed for official registration!).
- Description of the MR-Job (max. 500 words)
- Java classes (source code! - all with appropriate comments)



## Assignment 2 – Hadoop Ecosystem Tool

Choose a tool from the Hadoop Ecosystem

Present it in the lecture:

- What is the purpose of the tool?
- Give a brief overview of its architecture
- Examples of how it is used (Use cases, maybe in combination with other tools)
- Short Demo if possible

A list of possible tools is available under [doodle:http://doodle.com/nqpx35m8xxgqu5k8](http://doodle.com/nqpx35m8xxgqu5k8)

Each group should select one of them. None of the tools should be selected more than once! (First Come First Serve).

## Assignment 2 – Submission

- Data of Group members including: „Name, Matrikelnummer und Studiengang”.
- Description of your presentation (max. 500 words) **OR** presentation slides.

