

Concepts and Technologies for Distributed Systems and Big Data Processing

Guido Salvaneschi

People



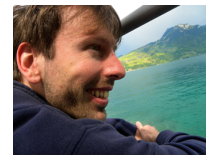
Guido Salvaneschi



Pascal Weisenburger



Matthias Eichholz



[...]

Motivation

- Distributed systems are ubiquitous
 - Emails, WWW, ...
- Area is quickly changing - sorry for the buzzwords :)
 - Big data analytics
 - Complex event processing/stream processing
 - Cloud computing
 - New programming models (streams, actors)
 - ...
- Yet, many concepts remain the same
 - Abstraction over low-level details
 - Fault tolerance
 - Performance: throughput, latency
 - Asynchronous communication
 - ...



Concepts and Technologies for Distributed Systems and Big Data Processing

- Provide an overview of recent development in distributed systems and Big Data processing
- Focus on concepts, not on technology.
 - Technology was different 5 years ago, will be different in 5 years
 - We will look at technology only as a way to better grasp the concepts
- Selected topics
 - Different speakers for each topic + guest lectures
 - Dr. Guido Salvaneschi, Dr. Michael Eichberg, Prof. Patrick Eugster, Dr. Alessandro Margara, Prof. Philipp Haller, ...

About this course

What this course **IS**

- An introduction on selected topics
- ...on distributed systems and big data processing
- A course about concepts... and a bit about technologies

What it is **NOT**

- It is not (only) a course about recent trends in Big Data.
- It is not a systematic course on distributed systems
 - Distributed Systems: Principles and Paradigms
Andrew S. Tanenbaum, Maarten van Steen
- It is not a tutorial on how to program big data systems

Tentative course schedule

*Big data and
complex event
processing*

April 12 - Intro, motivation
April 19 - Intro to big data, mapreduce
April 26 - hadoop, hdfs
May 4 - Complex event processing

May 10 Complex event processing

Date and room change!
11:40-13:20 CET
room S101/A03

*Architectures and
programming models
for distr. systems*

May 17 – Concurrent programming
May 24 - Futures, actors, streams
May 31 - Component-based architectures
June 7 - REST architectures

*Cloud computing and
advanced topics in
resource management*

June 14 - Spark streaming / Spark ?
June 21 - Geodistribution
June 28 - Security, Resource management
July 5 – DB Guest lecture ?
July 12 – Guest lecture/Exam preparation

July 26 - Exam



Subject to change!
See course website
for updates

Exercises

- No graded exercises
- Exercises will be provided after some lectures.
- Solutions discussed in the next lecture if needed

- Why
 - Get more confident on the topics of the lectures
 - Get a feeling the type of questions that can come up in the exam

- Examples
 - Read a paper
 - Answer questions based on the content of the lecture
 - Discuss a case study
 - Small coding exercises

Registration and Grading

- **Register in Tucan for the course!**
- Written exam – July 26
 - 90 mins
- Questions about the topics covered in the lectures
- Simple “programming” tasks
 - Understand a code snippet that is relevant for one of the covered topics
 - Complete a code snippet

Resources

- Website (slides, exercises)
 - <http://stg-tud.github.io/ctbd/>
 - This is the official place slides, exercises, updated schedule, etc.
- Forum
 - <https://www.fachschaft.informatik.tu-darmstadt.de/forum/viewforum.php?f=580>
 - Please use the forum for your questions.
Answers will remain as a reference for other people.
- Please ask assistants for questions that cannot be posted on the forum
 - But think if this is really the case :)