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
 - Asynchonous 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 topics, selected among the "most interesting" ones
 - Guest lectures

About this course

What this course IS

- An introduction on <u>selected</u> 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 <u>not a systematic</u> course on distributed systems
 - Distributed Systems: Principles and Paradigms Andrew S. Tanenbaum, Maarten van Steen
- It is <u>not a tutorial</u> on how to program big data systems

Tentative course schedule

Big data and complex event processing

Architectures and programming models for distr. systems April 21 - Intro, motivation
April 28 - Intro to big data, mapreduce
May 5 - hadoop, hdfs
May 12 - Futures, actors, streams
May 19 - Futures, actors, streams
May 26 - Complex Event Processing
June 2 - Spark
June 9 - Spark streaming
June 16 - Exam preparation

July TBA - 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

<u>Register in Tucan for the course!</u>

- Written exam July, date TBA
 - 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.tudarmstadt.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 :)