

Advanced Seminar Introduction SS 2014

Internet-Scale Distributed Systems

(open for Bachelor and Master students)

TUM Informatik XIII
Application and Middleware Systems

<https://msrg.in.tum.de/>

Andreas Veit

Outline

- Overview & Introduction
- Organization
- Topics
- Goals & Expectations
- Effort & Deliverables
- Marking
- Topic Allocation

Overview teaching SS 2014

Lectures

Event
Processing

Seminars

Agiles
Geschäfts-
prozess-
management

Internet-
scale
Distributed
Systems

Energy
Informatics

Modern
Hardware in
Big Data
Processing

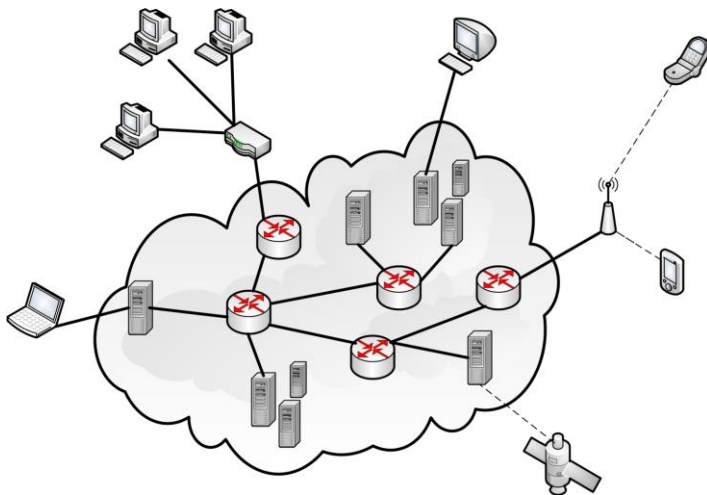
Pro Seminar

“H4ck3rs D3l1ght” (für Bachelor Studenten)

Lab Course (Praktikum)

Research on Distributed Systems & Middleware

Internet- scale Distributed Systems



Source: <https://www.google.com/about/datacenters/inside/locations/index.html>

Internet- scale Distributed Systems

We'll look at systems running at major internet players



Design and architecture of

- File systems
- Locking services
- Messaging systems
- Key-value stores
- Data analytics platforms
- Notification services

Outline

- Overview & Introduction
- **Organization**
- Topics
- Goals & Expectations
- Effort & Deliverables
- Marking
- Topic Allocation

Organization

- **Overview**
 - 4 ECTS Credits
 - In English language
 - Open for Bachelor and Master students

Organization

- Introduction, topic allocation and Kick-off
 - Today
 - One or, preferably **two** students per topic (depends on general interest and topic complexity). Students working in group will get bonus.
- Presentations
 - Every Wednesday during the time (14:00 – 15:30)
 - One or two presentation per session according to topic ordering
 - The slides of each presentation should be delivered to me, no later than **the day before the presentation's day**
- Written Report
 - To submit **two weeks** after your presentation

Outline

- Overview & Introduction
- Organization
- **Topics**
- Goals & Expectations
- Effort & Deliverables
- Marking
- Topic Allocation

Topics (1)

- 1) Review and present the main ideas underlying the **CAP** theorem along with its proof sketch, and explain the relation between CAP theory and **Inktomi** architecture (Cluster-Based Scalable Network Services).
- 2) Review and present the main ideas underlying the different **consistency models**, and explain their advantages and disadvantages from several perspectives.
- 3) Review and present the **Google File System** and the **Hadoop Distributed File System**, and also explain the similarities and differences between these two Internet-scale distributed systems from several perspectives.

Topics (2)

- 4) Review and present Yahoo's **Zookeeper** and Google's **Chubby**, and also explain the similarities and differences between these two lock services from several perspectives.
- 5) Review and present Google's **Thiafi** and LinkedIn's **Kafka**, and also explain the similarities and differences between these two Internet-scale distributed applications from several perspectives.
- 6) Review and present Google's **BigTable** storage and Amazon's **Dynamo** storage, and compare these two systems, and finally explore how CAP theory is followed in these systems.
- 7) Review and present Yahoo's **PNUTS** and Facebook's **Cassandra**, and compare these two systems, and finally explore how CAP theory is followed in these systems.
- 8) Review and present Google's **Spanner**, and also explain the similarities and differences to the traditional **NoSQL** approaches from several perspectives.

Topics (3)

- 9) Review and present the **Map-Reduce** computational model and **HadoopDB**, and then describe how and why HadoopDB takes advantage of Map-Reduce model in its implementation.
- 10) Review and present the **Spark** model, and also explain the similarities and differences to the **Hadoop-style Map-Reduce** from several perspectives.
- 11) Review and present Facebook's **Hive** and **Shark**, and also explain the similarities and differences between these two Internet-scale distributed applications from several perspectives.
- 12) Review and present the **MLBase** system, and then explain how and why MLBase takes advantage of the **Spark** model in its implementation?

Topics (4)

- 13) Review and present Google's **Megastore** and Google's **Dremel**, and compare these two Internet-scale distributed applications from several perspectives.
- 14) Review and present Google's **Percolator** and Google's **Dapper**, and compare these two Internet-scale distributed applications from several perspectives.

Outline

- Overview & Introduction
- Organization
- Topics
- **Goals & Expectations**
- Effort & Deliverables
- Marking
- Topic Allocation

Goals & Expectations

- **Goals**
 - Overview of research area
 - Improve presentation skills (i.e., soft skills)
 - Prepare for independent research
- **Expectations**
 - Genuine interest and active participation with questions & feedback, etc.
 - Deliverables: presentation, slides, written report
 - 30-45 minutes presentation & discussion
 - Required attendance & active participation

Outline

- Overview & Introduction
- Organization
- Topics
- Goals & Expectations
- **Effort & Deliverables**
- Marking
- Topic Allocation

Deliverables

- **Slides**
 - Not just lists of bullets
 - Do not copy figures & graphs (95% of slides must be your own design)
 - Show didactic plan
 - Copying and plagiarism will reduce mark considerably
 - **Submit slides no later than one day before the presentation's day**
- **Report**
 - Summarize your results (overview, comparison, other issues,...)
 - 5-page ACM proceedings style (template will be delivered)
 - **Submit the report by no more than 2 weeks after the presentation**

Outline

- Overview & Introduction
- Organization
- Topics
- Goals & Expectations
- Effort & Deliverables
- **Marking**
- Topic Allocation

Marking I - Criteria

- **Presentation (50%) - Well structured and formatted slides**
 - Presentation style (soft skills)
 - Scope of presentation (technical content)
 - Grading rubric
- **Report (50%): Classical paper structure**
 - Publishable quality (spelling, grammar)
 - Content/ contribution
 - Evaluation
 - Motivation, related work, conclusion

Marking II – Rubric for presentation

- Evaluation scale
 - Outstanding (5), excellent (4), – incomplete (1)
- Four mark categories
 - Presentation content (5 aspects) 25 points
 - Performance (5 aspects) 25 points
 - Visual aids (5 aspects) 25 points
 - General (5 aspects) 25 points

(max 100 points)

Seminar Rubric: **Agile Workflow Management and Execution** **WS '12**

Topic: _____

Date: _____

Presenter: _____

Judge: _____

Requirement	very poor (1)	poor (2)	fair (3)	good (4)	very good (5)	Comment
Presentation Content:						
Simplicity appropriate for audience						
Problem clearly stated and motivated						
Background and context of research demonstrated						
Results easily and clearly described						
Conclusions to the point, corresponding to problem						
Visual Aids:						
Clear, catchy slides, not overloaded						
Contribution of colors and font						
Legible (self-made) figures conveying results effectively						
Vivid examples and animations						
Overall homogeneity						
Performance:						
Smooth transitions from issue to issue						
Integration of the audience						
Voice and pacing						
Engagement and enthusiasm						
Body language, gestures						
General:						
Organization of information (logical ordering, structure)						
Ability to answer questions (defending the presentation)						
Adherence to time limit (effective use of time)						
Minimum redundancy						
Slides support talk						

Feedback

Over-all: Points: _____ Mark: _____

Outline

- Overview & Introduction
- Organization
- Topics
- Goals & Expectations
- Effort & Deliverables
- Marking
- **Topic Allocation**

Topic Allocation Discussion

Via this: [Google Form](#)

Until next Wednesday February 12th

Preliminary allocation and schedule send out on Feb 19th

Confirm your selected topic and date by March 12th

Otherwise, it is assumed that you are not going to take this course and your topic will be unallocated.

Seminar Organization

- **Where** Room 01.06.011
- **When** Wednesdays 14:00 – 15:30
- **Start** Wednesday April 09th 2014
- **Contact** **Andreas Veit**
 - Email a.veil(at)in.tum.de
 - Room 01.06.039
 - Phone 089 / 289 18485

Questions ?