

Data Processing Architectures for New Hardware Platforms

Introduction

Gustavo Alonso, Andrew Baumann, Donald Kossmann,
Timothy Roscoe, Nesime Tatbul, Jens Teubner



Overview

- ▶ Hardware architecture is changing rapidly
 - ▶ Multi-core and multi-threading
 - ▶ Memory bottleneck:
 - ▶ Increased dependence on caches
 - ▶ Cache-locality more important
 - ▶ NUMA
 - ▶ New storage technologies
- ▶ How should data processing systems change?
 - ▶ Do the classic structures and optimisations still make sense?
- ▶ This seminar: discuss and critique recent work in this area
 - ▶ Wide range: from database systems, to storage and OS fields
- ▶ Our sub-agenda: exploring ideas for DB and OS co-design

Format

- ▶ Two topics per week
 - ▶ Most topics have one paper
 - ▶ In other cases there is another paper provided as background reading
- ▶ Everyone reads all the papers
 - ▶ Emails short summary in advance
- ▶ Allocated student presents each paper
 - ▶ Approx. 20 minutes plus questions
 - ▶ Everyone discusses
- ▶ Written report (~3 pages) on the topic of your presentation due by year end (date TBA)

Papers

- ▶ All from recent major conferences/journals
- ▶ Read positively, but critically
 - ▶ These are all good papers
 - ▶ What should we take away, and what should we discard?

Pick a topic

- ▶ Topics for the first two seminars available for allocation now
- ▶ Otherwise email Andrew <[andrewb@inf](mailto:andrewb@inf.ethz.ch)> with your preferences by Thursday
- ▶ We'll let you know

Reports (just the speakers)

- ▶ Due by the end of the year (date TBA)
 - ▶ No harm in getting them done early!
- ▶ Approximately 3 pages on the system and the paper(s)
- ▶ Summarise your talk, don't just paraphrase the paper(s)
- ▶ Include relevant material from the discussion

Talks

- ▶ Aim for 20–25 minutes speaking time
- ▶ Slides and/or whiteboarding
- ▶ Pick out important aspects of paper and present
 - ▶ Remember, everyone's read them (in theory)
 - ▶ Don't just regurgitate the results
 - ▶ Feel free to pose your own questions!

Consultations

- ▶ Andrew and Jens are available for consultation
 - ▶ Getting advice on your presentation is encouraged!
- ▶ Arrange a meeting time in advance (eg. per email)
 - ▶ Say what you want to talk about
 - ▶ Don't just turn up!

Questions after the talks

- ▶ Ask away!
- ▶ Everyone should participate in the discussion
 - ▶ Part of the grade will be allocated for participation
- ▶ Wider/general discussion is fine
- ▶ Think about what you might ask when writing the summary

Summaries (everyone)

- ▶ 2–3 paragraphs
 - ▶ What are the important ideas?
 - ▶ 1–3 strengths of the work
 - ▶ 1–3 weaknesses of the work
 - ▶ 2–3 questions about the work, for the discussion
- ▶ Due the preceding Monday night
- ▶ Email to both Andrew <andrewb@inf> and Jens <jens.teubner@inf>

Evaluation Criteria

- ▶ Quality of talk
 - ▶ Motivated
 - ▶ Clearly explained
 - ▶ Knowledge transfer
 - ▶ Difficulty
 - ▶ Prior knowledge
 - ▶ Structure
 - ▶ Encouraged participation
 - ▶ Media
- ▶ Quality of report
- ▶ Quality of summaries
- ▶ Participation, questions, etc.

Papers

1. SEDA
2. Cache-conscious radix-decluster projections
3. Optimistic intra-transaction parallelism on CMP
4. Improving DB performance on SMT processors
5. MonetDB/X100
6. Super-scalar RAM-CPU cache compression
7. On multidimensional data and modern disks
8. Rethink the sync
9. Cooperative scans
10. Clock scan
11. QPipe
12. Corey

