Seminar: Modern Database Systems

Kickoff Meeting

Prof. Dr. Jana Giceva
Per Fuchs, M.Sc.
February 04, 2021
Overview

Weekly Meeting

- Monday, 14:00 - 16:00, starting April 26th, 2021
- in this BBB room
- 2 presentations per meeting
- There will be an attendance log and participation is part of your grade

Required Work

- Seminar paper (≤ 5 pages) - 60% of the grade
- Including a short description of your own idea for future research and the expected challenges
- Presentation (20 minutes + 10 minutes discussion) - 30% of the grade
- Moderate one discussion (act as the "devil's advocate", you should pair up for this) - 10% of the grade
Organization

Registration through matching system
- Write an email to per.fuchs@cs.tum.edu if you are interested
  - Subject should be: [DBSeminar] Kickoff Meeting
  - Emails will be filtered based on this subject
- Register for the seminar on https://matching.in.tum.de!

After matching: choose a paper to present
1. Email 3 preferences soon after the matching has been finalized (preferences considered FCFS)
2. Check in when rough structure is planned
3. Check in when final draft is ready

Due Dates
- Structure: ca. 4 weeks prior to presentation date
- Presentation slides: 1 week prior to presentation date
- Seminar paper final handin: 2 weeks after presentation date
Topics List

Block 1: Distributed Systems Challenges

- A Critique of the CAP Theorem (arXiv 2015)
- Paxos made simple (SIGACT 2001)
- In search of an understandable consensus algorithm (Raft, Usenix 2014)
Topics List

Block 2: Consistency

- Consistency, Availability, and Convergence – Mahajan et al. (TR from UT Austin, 2011)
- Don’t settle for Eventual: Scalable Causal Consistency for Wide-Area Storage with COPS – Lloyd et al. (SOSP’11)
- A comprehensive study of Convergent and Commutative Replicated Data Types – Shapiro et al. (2011)
Topics List

Block 3: Database Systems

- Anna: a KVS for Any Scale – Wu et al. (ICDE’18)
- Cloud-Native Database Systems at Alibaba: Opportunities and Challenges (VLDB 2019)
- Building an Elastic Query Engine on Disaggregated Storage. - Vuppalapati et al. (NSDI 2020)
- POLARIS: Distributed SQL Engine in Azure Synapse. - Aguilar-Saborit (VLDB 2020)
- CockroachDB: The Resilient Geo-Distributed SQL Database (SIGMOD 2020)
Topics List

Block 4: Transactions in the Cloud

- Obladi: Oblivious Serializable Transactions in the Cloud - Crooks et al. (OSDI’18)
- FaSST: Fast, Scalable and Simple Distributed Transactions with Two-Sided (RDMA) Datagram RPCs - Kalia et al. (OSDI’16)
- Highly Available Transactions: Virtues and Limitations – Bailis et al. (VLDB’14)
- No compromises: distributed transactions with consistency, availability, and performance - Dragojevic et al (SIGMOD’15)
Topics List

Block 5: Serverless Data Processing

• Shuffling, Fast and Slow: Scalable Analytics on Serverless Infrastructure - Pu et al. (NSDI’19)
• Lambada: Interactive Data Analytics on Cold Data Using Serverless Cloud Infrastructure - Mueller et al. (SIGMOD’20)
• Starling: A Scalable Query Engine on Cloud Functions Matthew - Perron (SIGMOD’20)
• Magnet: push-based shuffle service for large-scale data processing (VLDB 2020)
What are seminars about?

1. **Our papers are chosen to get a good overview on Cloud-Based Data Processing**
   - From the background reading of the lecture last year
   - As base for my further PhD

2. **Learn how to present a novel idea to an audience in an understandable and engaging way**
   - Summarize a years worth of work in 20 minutes
   - Present new ideas instead well-established topics
   - Work is to big parts judged based on presenting skills

3. **Learn about academic writing**
   - Important for your thesis
How to read a paper?

Read a paper at least three times

First pass: get the general idea
- 5 - 10 minutes
- Read abstract, introduction, conclusions, (sub-)section headings

Second pass: understand the content
- Roughly 1 hour
- Read the full paper, ignore details (e.g. proofs, etc.)
- Find key points, take notes, check figures carefully to understand them
- Mark references which should help you in understanding the paper

Third pass: understand in depth
- 4-5 hours
- Fully understand everything, attention to detail (check related work)
- Try to virtually reimplement the paper
- Question everything
- Generate ideas for your work
Some model questions

What is the research question that the paper addresses?
• What is the motivation? Is it relevant? What is the impact if it is solved?

What are the contributions?
• How do they build on previous work? Is it something new? Can it be generalized?

What do I learn by reading it?
• For instance, even a good summary of related work can be worth a lot!

How are the results substantiated?
• Experiments, proofs, benchmarks, etc. Is the evaluation thorough?

What are the conclusions and broader impact?
• What can be built on top of it? What is an interesting follow-up / future work?

Tip: you should be able to answer all of these questions before you present a paper!
A good (research) talk

Is centered around the audience (not you)
  • Teaches, engages, provokes and excites the listener

Provides intuitions to the audience

Should make them want to read the paper
  • But not because they did not understand you

It should not
  • Tell them all the technical details
  • Cover everything you know about the topic
  • Impress the audience how smart you are
What to put in your talk?

**Motivation (20%)**
- Why should I listen to the talk?
- What is the problem?
- Is it an interesting problem?
- Is the proposed solution worthwhile?

**The key idea (80%)**
- Identify the key idea
- Be specific - do not leave it up to the audience to figure it out
- Be **very** specific - say "if you remember nothing else, remember this"
- Organize your talk around this specific goal. Remove everything that is irrelevant.

**Avoid shallow overviews**
- Cut to the technical "meat" even if it covers only a part of the paper

**Use examples!**
- To motivate work, to convey basic intuition, to illustrate the idea, etc.
What to omit?

Do not present talk outlines
  - It is not informative, however, a good talk has a storyline

Do not present excessive related work
  - But mention it in your slides, acknowledge pre-cursors, praise the opposition

Do not present too many technical details
  - The audience may find it difficult to follow
  - Put them in the back-up slides in case somebody asks

Do not exaggerate with animations
  - Animations are good but can be also distracting

Do not clutter your slides with graphics
Pointers to further reading

**How to read a paper:**
- S. Keshav - How to read a paper?
- Philip W. L. Fong - How to read a CS research paper?

**How to give a good (technical/research) talk:**
- Markus Puschel - How to give a good technical presentation?
- Simon Peyton Jones - How to give a great research talk?

**Academic writing:**
- Will be provided on Moodle when the course starts.