Kick-Off

Course Goals

Primary goals:

- Learn how to write a scientific paper
- Learn how to give a scientific talk
- Structure ideas from multiple research papers

Secondary goals:

- Practice to read scientific papers
- Understand modern cloud database system architectures
- Have interesting discussions with your peers

What you get out of this course

In this course you will get:

- At the beginning of the seminar we will give three introduction lectures:
  - Introduction to distributed cloud databases
  - How to write a report
  - How to give a presentation
- Two peer reviews for your paper
- Presentation feedback from your peers

Prerequisites

Hard prerequisite:

- Introduction to Fundamentals of Databases (IN0008) or equivalent

Beneficial previous knowledge:

- Query Optimization (IN2219)
- Database Systems on Modern CPU Architectures (IN2118)
- Most courses from Prof. Neumann, Prof. Giceva, Prof. Kemper, and Prof. Leis

Deliverables

During the course each of you will create:

- A report proposal that summarizes your reports topic and focus ($<$1 page)
- A research report **draft** (5-8 pages + references)
- 2 peer reviews for your fellow students ($<$1 page each)
- A research report (5-8 pages + references)
- A pre-session protocol that shows your preparation for the presentation. ($<$1 page)
- A presentation (13-15 minutes)
- Meaningful contribution to the panel discussion in your presentations session

Grading

Rough estimate of grading contributions:

- $\approx 45\%$ Report
- $\approx 25\%$ Presentation
- $\approx 10\%$ Report Proposal
- $\approx 10\%$ Peer reviews
- $\approx 10\%$ Pre-session protocol and panel discussion

This is subject to change!

Topics (1)

Preliminary list of topics:

- **Neojoin:** How to do distributed joins fast: NeoJoins and evolution
- **Flowjoin:** How to handle skew in distributed joins
- **Snowflake:** A distributed OLAP cloud Unicorn
- **Firebolt:** Building a good system fast
- **Redshift:** How Amazon scales for larger datasets
- **Polaris:** How Microsoft scales virtually infinitely
- **Trino:** How Meta analyzes large data from many sources
- **Velox:** Meta's faster Trino
- **Lambada:** Completely serverless
- **Client-Server:** Hand off computation to clients

Topics (2)

- **Anyblob:** Single node storage decoupling
- **Microsoft's Optimizers:** Early ideas SQL server PDW and heuristics
- **MemSQL:** Query Optimzation is hard
- **Vertica:** uses vertical fragmentation
- **Eigen:** Alibaba's cluster scheduling
- **Key Recommendation:** for better schemata
- **EdgeFrame:** Worst-Case Optimal Joins and Graph Pattern Matching
- **Scyper:** single node + log replication is fast
- **Socrates:** Scaling OLTP
- **Noria:** fast analytics in OLTP systems

Your ideas for related topics are very welcome!

Timeline

Preliminary timeline:

- Mo 28.04.2025 introduction lecture 01
- Mo 05.05.2025 introduction lecture 02 | **submit** topic preferences
- Mo 12.05.2025 introduction lecture 03
- Mo 19.05.2025 ∅ | **submit** report proposal
- Mo 26.05.2025 ∅
- Mo 02.06.2025 ∅
- ~~Mo 09.06.2025~~ Pfingstferien
- Mo 16.06.2025 ∅ | **submit** report draft
- Mo 23.06.2025 ∅
- Mo 30.06.2025 presentation session 01 | **submit** peer reviews
- Mo 07.07.2025 presentation session 02
- Mo 14.07.2025 presentation session 03 | **submit** final report
- Mo 21.07.2025 presentation session 04
- **Attendance to all sessions in presence is mandatory**

Organization

- Attendance to all sessions in presence is mandatory
- You have to write your paper in LaTeX using our template
- Register on our gitlab server with your TUM email to get access
- Communication will take place using our Mattermost instance

Matching

Register for the course through the matching platform (https://matching.in.tum.de/)

Contact

- Website: http://db.in.tum.de/teaching/ss25/seminarModernDatabaseSystems
- Maximilian Rieger: riegerm@in.tum.de