

Seminar: Next Generation Database Systems

Kick-off Meeting

Michalis Georgoulakis, Tobias Götz

Garching, July 9th, 2026



Meet the TAs: Who am I?

Michalis Georgoulakis

Email: michalis.georgoulakis@tum.de

Office Number: 02.11.060

Webpage: <https://db.in.tum.de/people/sites/georgoulakis/>

PhD candidate at the Chair for Database Systems

M. Eng National Technical University of Athens

Cloud-based analytics, Caching architectures, Query optimization, Scheduling



Meet the TAs: Who is Tobi?

Tobias Goetz

Email: goetzt@in.tum.de

Office Number: 02.11.055

Webpage: <https://db.in.tum.de/people/sites/goetz/>

PhD candidate at the Chair for Database Systems

MSc Technical University of Munich

Query Processing on OTFs and Elastic Compute

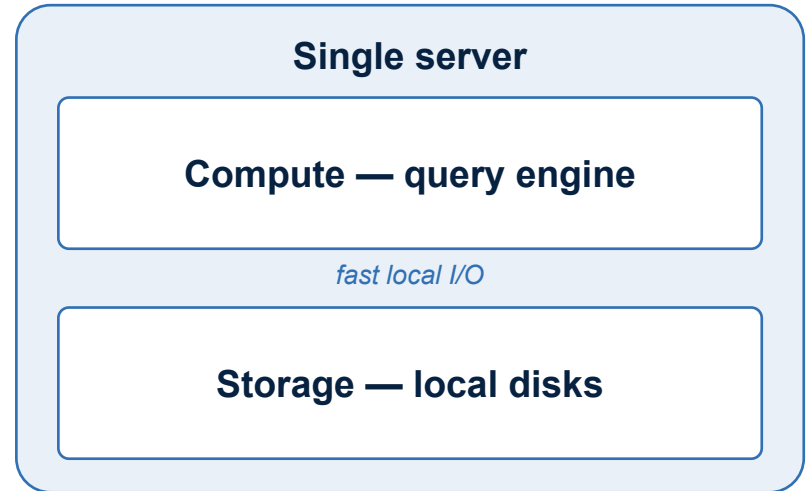


Motivation: Traditional Database Architecture

For decades, the whole system lived on one machine.

- Query engine, memory and disks sat side by side.
- The system knew its hardware, and data sat right next to the compute using it.
- This assumption shaped how we built systems.

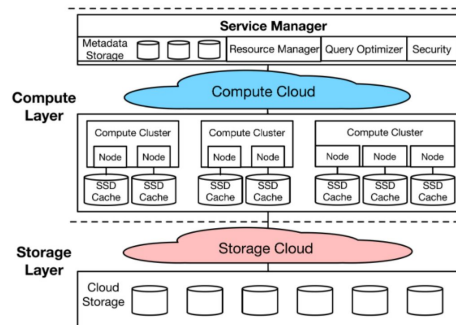
The cloud broke this apart.



Storage and Compute Pulled Apart

What the cloud gives you

- Compute runs on elastic, stateless nodes.
- Data lives in cheap, shared object storage (S3, GCS, ...).
- Scale each independently; pay only for what you use.



Src: Li et al. Cloud-Native Databases, VLDB'22

New problems

Caching & pushdown across the network

ACID transactions over object stores

Keeping the data lake open & interoperable

Multi-cloud operations

Elastic, serverless execution

... and now the Hardware is shifting too

Hardware accelerators

GPUs, FPGAs and custom silicon take over query operators that CPUs handle too slowly.

→ **CoddSpeed — Microsoft Fabric**

Disaggregated memory (CXL)

Memory is pooled across machines over CXL, no longer trapped inside one server.

→ **Pond — CXL-based memory pooling**

- **How do we build systems under these new assumptions?**
- **How do we get the best out of the latest hardware?**

You will get to know :)

Structure and Grading

Preliminary distribution for the final grade:

- Report (~50%)
- Presentation (~30%)
- Moderation and Discussion (~20%)

Attendance during all presentations and discussions is mandatory to receive a passing grade!

In case of an urgent event, send us an email beforehand.

Do not use LLMs to write your submission!

The Role of Supervisors

Each paper is supervised by a specific supervisor

Individual submissions will be graded by the respective supervisor of the topic

- Can help you to **understand** the contents of the paper
- Might **discuss** some details of a paper with you
- Provides **feedback** to your submissions
- Will be **present during your presentation** and **join the discussion**
- Will **grade** your individual submissions

Besides us, there may be some additional Ph.D. students from our group who will participate as supervisors.

Report

Should contain:

1. A summary of the paper
2. Review of the paper (see discussion)
3. ~1 page about your own research idea
 - a. Problem statement
 - b. Expected impact
 - c. Research questions
 - d. Microbenchmarks & Approach

The report should be **≤ 5 pages** and formatted as for SIGMOD (template on the course website)

Writing hints will be published and discussed after the first meeting

The screenshot shows the TUM website for the 'Seminar - Modern Database Systems' course. The page is in German and English. The main content area includes the following sections:

- Seminar - Modern Database Systems**
 - Prof. Dr. Jana Giceva
 - Simon Ellmann, M.Sc.
 - Tobias Goetz, M.Sc.
 - Adrian Riedl, M.Sc.
 - Stefan Lehner, M.Sc.
 - Altan Birler, M.Sc.
 - Michail Georgoulakis Misegianis, M.Eng.
- Please contact us using** mds@mailkemper.in.tum.de.
- Meetings** take place on Mondays (2pm-4pm) in 02.09.014
- Content**

In this seminar, we study the challenges of modern database systems. We discuss topics along very recent publications in seven sessions.
- Prerequisites**

This seminar is geared towards the top students in database systems.

 - Fundamentals of Databases* (Grundlagen Datenbanken, GDB) or similar course.
 - Additional database implementation courses are highly recommended.
- Links**
 - presentation guide
 - Latex/Word template
 - How to write a paper
 - Writing tips
- Dates and Deadlines**
 - Preliminary discussion:**
 - Tuesday, February 7, 2023: 2pm till 4pm in this BBB room

A red arrow points to the 'Links' section.

Presentation

15-20 minutes; 20 minutes is a strict upper bound

Onsite (days and room TBA)

Summarizes the paper!

- Key contributions
- Ideas and concepts
- Experiments and results

Does not contain:

- Your report
- Your own research idea

A Good Presentation...

- ... should define a problem
- ... state the contributions
- ... use examples, figures, or more to explain the idea
- ... explain the most important experiments and their results
- ... finish with its main takeaways

Moderation

Minimum 10 minutes after the respective presentation

You will moderate a different paper than you present

Acknowledge questions

- Order all questions and call all questioners
- Help the presenter and the participants to find an answer

Prepare some **discussion aspects** (e.g. might contain aspects from the VLDB review process)

- Novelty
- Significance (determine the impact of the paper)
- Depth, quality, and chosen experiments
- Presentation of the paper's contents



Discussion

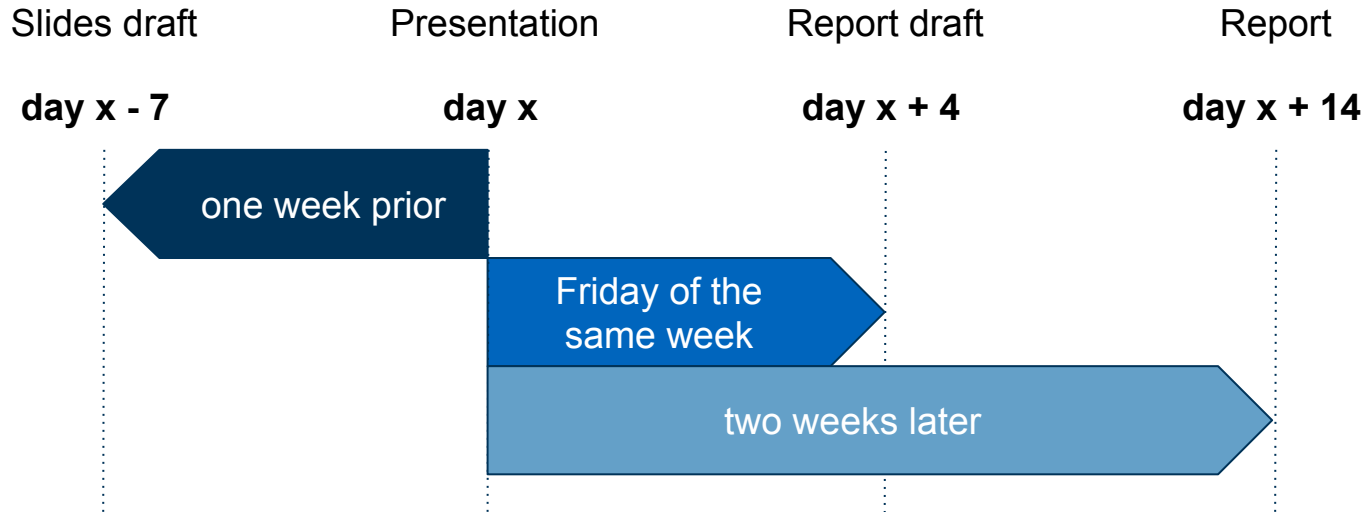
Everyone in this seminar is expected to read a paper before its presentation

- You can take notes of your ideas while you read
- Your thoughts, ideas, alternative designs, or input from your previous experience are valuable!
- Even if something is unclear (could be discussed as well)

Do not hesitate to ask questions to all participants or start a new discussion

Your participation in the discussions is part of your final grade

Deadlines



Sample of Papers

- LiquidCache: Efficient Pushdown Caching for Cloud-Native Data Analytics
- CoddSpeed: Hardware Accelerated Query Processing in Microsoft Fabric
- Skyrise: Exploiting Serverless Cloud Infrastructure for Elastic Data Processing
- Delta Lake: High-Performance ACID Table Storage over Cloud Object Stores
- Pond: CXL-Based Memory Pooling Systems for Cloud Platforms
- Active Data Lakes: Regaining Physical Data Independence Without Losing Interoperability
- BigLake: BigQuery's Evolution toward a Multi-Cloud Lakehouse

Communication and Material

Officially via e-mail (michalis.georgoulakis@tum.de, goetzt@in.tum.de)

Use Mattermost for questions

All materials will be available on GitLab

Next Steps

When registration opens: Register your priorities via Matching System

The matching results will be published at the Matching System.

12.10.2026 – 05.02.2027: Teaching Period

Questions?

