Seminar: Next-Generation Data Systems

Session 3: Topics and Academic Writing

Michalis Georgoulakis, Tobias Götz

Garching, October 28th, 2026



Today

Your topics and the schedule

- Who presents and moderates which paper, and when
- Deadlines and feedback meetings

Academic writing

- Your report: what goes in, and how to write it
- Clear sentences and practical tips

Topic Assignment (1)

Date	Theme	Paper	Presenter	Moderator	Supervisor
02.12.2026	**Disaggregated cloud OLTP**	**Amazon Aurora**: Design Considerations for High Throughput Cloud-Native Relational Databases	Miguel	*Quirin*	*Tobias*
		Socrates: The New SQL Server in the Cloud	Van Trang	*Maximilian*	*Michalis*
09.12.2026	**Going distributed and multi-region**	**Aurora DSQL**: Scalable, Multi-Region OLTP	Mahdi	*Julian*	*Michalis*
		SunStorm: Geographically Distributed Transactions over Aurora-Style Systems	Khan An	*Samuel*	*Michalis*
16.12.2026	**Object storage and serverless compute**	**AnyBlob:** Exploiting Cloud Object Storage for High-Performance Analytics	Nourallah	*Benedikt*	*Tobias*
		Skyrise: Exploiting Serverless Cloud Infrastructure for Elastic Data Processing	Batuhan	*Ahmet*	*Michalis*
13.01.2027	**Lakehouse tables and data skipping**	**Delta Lake:** High-Performance ACID Table Storage over Cloud Object Stores	Zhenhan	*Yurii*	*Tobias*
		Pruning in Snowflake: Working Smarter, Not Harder	David	*Miguel*	*Tobias*

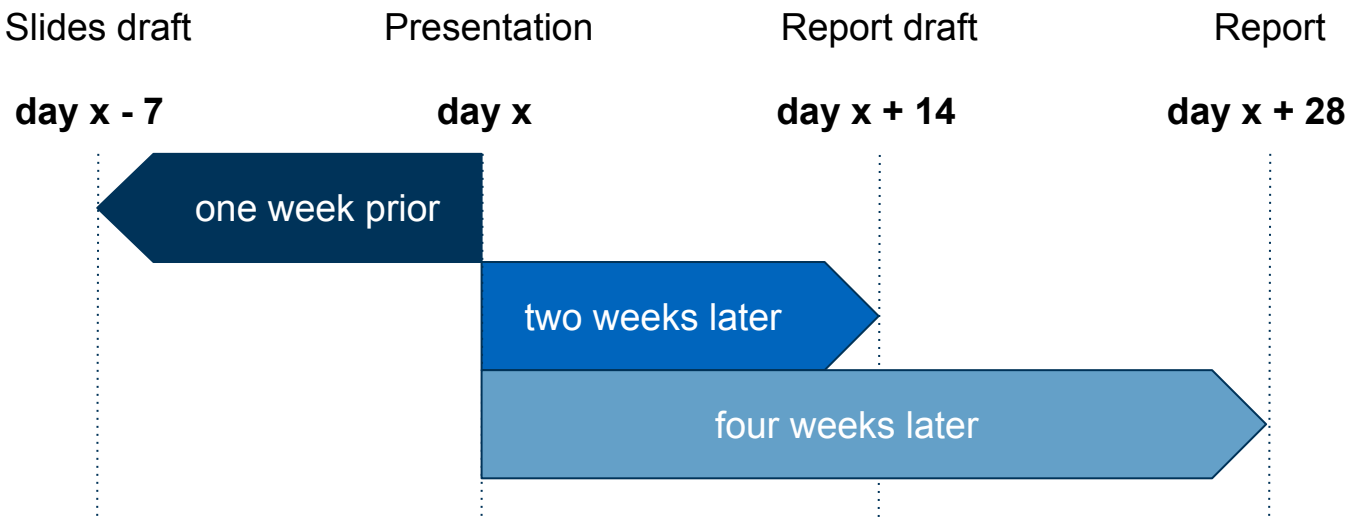
Topic Assignment (2)

Date	Theme	Paper	Presenter	Moderator	Supervisor
20.01.2027	**Computation pushdown**	**FlexPushdownDB**: Hybrid Pushdown and Caching in a Cloud DBMS	Maximilian	*Khan An*	*Tobias*
		LiquidCache: Efficient Pushdown Caching for Cloud-Native Data Analytics	Quirin	*Mahdi*	*Tobias*
27.01.2027	**Memory pressure: pooling and spilling**	High-Performance Query Processing with NVMe Arrays: **Spilling** without Killing Performance	Julian	*Nourallah*	*Tobias*
		CHARM: Chiplet Heterogeneity-Aware Runtime Mapping System	Benedikt	*Zhenhan*	*Michalis*
		Pond: CXL-Based Memory Pooling Systems for Cloud Platforms	Samuel	*Batuhan*	*Michalis*
03.02.2027	**Hardware acceleration**	**CoddSpeed:** Hardware Accelerated Query Processing in Microsoft Fabric	Ahmet	*David*	*Michalis*
		Sirius: Rethinking Data Processing in the GPU Era	Yurii	*Van Trang*	*Mykola*

Deadlines - Overview

Presentation day	Slides Draft	Report Draft	Report Deadline
02.12.2026	25.11.2026	16.12.2026	13.01.2027
09.12.2026	02.12.2026	23.12.2026	13.01.2027
16.12.2026	09.12.2026	13.01.2027	27.01.2027
13.01.2027	16.12.2026	27.01.2027	10.02.2027
20.01.2027	13.01.2027	03.02.2027	17.02.2027
27.01.2027	20.01.2027	10.02.2027	25.02.2027
03.02.2027	27.01.2027	17.02.2027	03.03.2027

Deadlines



Send all submissions **to your supervisor via e-mail.**

Exact dates, adjusted around Christmas: see the overview on the next slide.

Feedback Meetings

Presentation

- Send draft slides 7 days before your talk
- Meeting with your supervisor ~5 days before: explain your structure, what you included

Report

- Send your report draft ~2 weeks after your talk; meeting with your supervisor in the following week
- Final report ~4 weeks after your talk

Both meetings: mandatory, 30 minutes, in person in Garching (remote in exceptional cases).

Your Next Steps

1. **Read your paper**: pass 2 now, passes 3+ before you build your slides
2. **Contact your supervisor**: introduce yourself and fix the feedback meeting dates
3. **Read the paper you moderate**: and collect discussion points
4. **Mark your deadlines**: slides draft 7 days before your talk

Questions on Topics and Schedule?



Writing Is Thinking

“If you’re thinking without writing, you only think you’re thinking.”

Leslie Lamport

“Writing is thinking. To write well is to think clearly. That’s why it’s so hard.”

David McCullough

Unclear text usually means unclear thinking: fix the thinking first.

Your Report

At most 4 pages excluding references, SIGMOD format (template on the course website)

Summary · 1 p.	Critical review · 2 p.	Research idea · 1 p.
problem, key idea, results	your assessment, including points from the discussion	problem, impact, research questions, approach

Not a retelling of the paper: most of the report is your own thinking.

Know Your Reader

- **Your reader is your supervisor:** they know the paper well
- **So don’t retell the paper:** show your understanding and your judgment
- **Experts notice vague claims:** be precise, and back every claim
- **Your reader cannot ask you:** the text has to work on its own

If a reader misunderstands, don’t explain it to them: fix the text.

The Writing Process in Academia

```
def write():
    draft = write_draft()  # fast
    while quality(draft) < goal:
        draft = revise(draft)  # most time
    return draft
```

Algorithm from Viktor Leis, Writing for Computer Science (TUM)

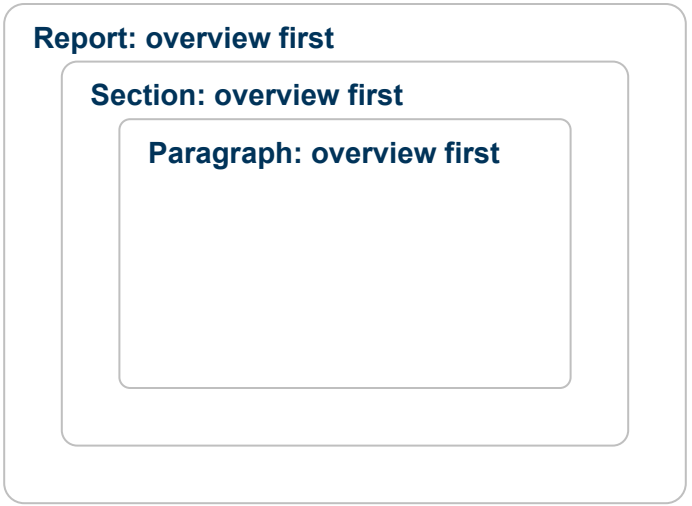
In practice

- Draft fast: perfect is the enemy of good
- Revise in short daily sessions
- Get feedback: the first reading is the most useful
- Kill your darlings: cut what doesn’t serve the point

Most of your time goes into revising, not into writing.

Structure: Top-Down

- **Big picture first,** then the details: in the report, each section, each paragraph
- **Open each section** with one or two sentences: where are we, what comes next
- **One paragraph, one point,** stated in its first (or last) sentence
- **Readers reduce your text to one key idea:** decide which one it is



The “onion principle”: write in layers, adding detail with each layer

Writing the Summary

- **What problem, why it matters,** the key idea, and the main results
- **In your own words:** not section by section
- **Concrete motivation** beats clichés

Before “In recent years, data volumes have grown rapidly, and the cloud has become a hot topic.”

After “When a database replicates every write across data centers, a conventional engine needs 7.4 network I/Os per transaction. Aurora needs fewer than one.”

Writing the Critical Review

- **Take a position** and argue for it; don't hedge to sound scientific
- **Back every claim:** a section, a figure, a number
- **Respond** to the points raised in the discussion
- **Be fair:** what works, what doesn't, and under which assumptions

Before "The evaluation is not very convincing."

After "The evaluation uses one dataset that fits in memory (Section 6), so it does not show what happens when data spills to SSD."

Your Research Idea

1. **Problem statement**: what is missing or broken, ideally a weakness from your review
2. **Expected impact**: who benefits, and by how much
3. **Research questions**: two or three, each one answerable
4. **Approach**: start from a simple baseline, then say what makes it hard

One page: a sketch, not a finished paper.

Characters and Actions

Subjects name the characters; verbs name their actions.

Before “The utilization of a buffer pool by the system enables a reduction of disk accesses.”

After “The system caches pages in a buffer pool, so it reads less from disk.”

Before “An evaluation of the approach on TPC-H was performed by the authors.”

After “The authors evaluate the approach on TPC-H.”

Hidden verbs: utilization → use, evaluation → evaluate, reduction → reduce · Joseph Williams, Style: Toward Clarity and Grace

Old Before New

Start a sentence with what the reader already knows; end it with what is new.

Before “Aurora sends only redo log records over the network. A quorum of four out of six storage nodes must acknowledge each write. Six copies in three availability zones store the data.”

After “Aurora sends only redo log records over the network. These records go to six storage nodes in three availability zones. A write is durable once four of the six nodes acknowledge it.”

The end of a sentence is its stress position: put the important part there.

One Term per Concept

- **Synonyms confuse:** readers assume different words mean different things
- **Introduce a term once:** in italics, and spell out acronyms
- **Name the ideas you talk about often:** a name makes them easy to refer to

Before “Delta Lake stores a table as Parquet files. Each object is immutable, so an update writes a new chunk.”

After “Delta Lake stores a table as Parquet files. Each file is immutable, so an update writes a new file.”

Cut Words

Wordy	Short
due to the fact that	because
in order to	to
is able to	can
prior to	before
it is possible that	may
a large number of	many
does not have	lacks

Example from Joseph Williams, Style: Toward Clarity and Grace

Before “In my personal opinion, we must listen to and think over in a punctilious manner each and every suggestion that is offered to us.”

After “We must consider each suggestion carefully.”

Don’t write what the reader can infer.

Common Mistakes (Not Only) by German Speakers

Instead of	Write
This approach allows to reduce overhead.	This approach allows us to reduce overhead.
There are less than five options.	There are fewer than five options.
We setup the server.	We set up the server. (The setup is …)
We proof that …	We prove that …
X is faster than Y. This is not well known.	X is faster than Y. This fact is not well known.
It's fast and doesn't block.	It is fast and does not block.

From Viktor Leis, Writing for Computer Science (TUM)

Figures, Tables and Citations

Figures

- Referenced in the text
- Caption says what to see
- Fonts as large as the text
- From the paper? Say so: “Figure from [1]”

Tables

- Graphs usually beat tables
- Only significant digits
- Use booktabs in LaTeX

Citations

- Invisible: “Snowflake skips partitions using min/max metadata [1]”, not “In [1], it was shown …”
- Copy paste BibTeX from DBLP

Tools and Final Touches

Tools

- LaTeX with the SIGMOD template, e.g., on Overleaf
- One sentence per line: easy to move, easy to diff
- Git: commit often

Final touches

- Spell check, LaTeX and BibTeX warnings
- Every figure referenced, every reference checked
- For paper submissions: Fix lonely last lines and bad page breaks last

Small errors signal a lack of care.

Revising

1. **Read it aloud**: you hear what your eyes skip
2. **Print it and use a pen**: a new medium gives a new perspective
3. **Jump to a random section**: and read it without the context
4. **Ask a friend to read it once**: then fix the text instead of explaining it
5. **Spell check last**: and once more after the last change

Revise, revise, revise.

LLMs and Your Report

Allowed

- Understanding background concepts
- Polishing the language of text you wrote
- “Find grammatical errors in this paragraph”
- “Flag overclaims and inconsistent terms”

Not allowed, and risky

- Generating your report text or your position
- Specifics are often wrong: citations, numbers, who said what
- Looks great today, reads verbose tomorrow

You own every claim, number and citation in your report.

Further Reading

- **Joseph Williams**: Style: Toward Clarity and Grace
- **Steven Pinker**: The Sense of Style
- **Justin Zobel**: Writing for Computer Science
- **Viktor Leis**: Writing for Computer Science (TUM course)
- **Larry McEnerney**: The Craft of Writing Effectively (talk, online)

Takeaways

1. **Writing is thinking**: unclear text often means unclear ideas
2. **Most of it is yours**: the review and the research idea carry the report
3. **Draft fast, revise most**: and let someone read it once
4. **Clear sentences**: characters as subjects, actions as verbs, old before new
5. **No waste**: one term per concept, no words the reader can infer

Next: preparation weeks (04.–18.11), the first talks are on 02.12