UV Distributed Information Management

Introduction

Daniel Kocher

Salzburg, March 15, 2023

Department of Computer Science
University of Salzburg

database
research group

PARIS
LODRON
UNIVERSITY
SALZBURG

Agenda

Today's Agenda

1. Introduction
2. Logistics
3. Outlook
4. Discussion

Who am I? Daniel Kocher, Research Assistant, Computational Systems Group

Database Research Group (summer 2019).

Contact & Lecture Policies

Contact (in this order):

1. **Lectures:**
   - Group 1: Mondays, 10:00 - 12:00 pm CET.
   - Group 2: Wednesdays, 01:00 - 02:30 pm CET.

2. **Slack:** https://dbteaching.slack.com/ (**create** an account with the stud email).

3. **Email:** dkocher@cs.sbg.ac.at (as a last resort).

Please **interrupt me** immediately if

- you have **troubles understanding** what I am talking about,
- you have a **question** related to the current topic, or
- there is an **error on my slides**[1] (0.5 bonus points/error; max. 5 per student).

---

[1]No punishment if it is not an error. Typos do not count (unless relevant), but please notify me anyways.

Why You Should Take This Course

The amount of **information** is growing rapidly and **needs to be managed.**

Many systems for different scenarios have been developed.

You must be able to **choose the proper system** for your use case.

Topics

Data Management:

- General-purpose and specialized database management systems.

- Declarative query processing.

- Relational and non-relational logical models.

- Workloads and challenges.

- Transaction models (ACID vs. BASE).

- CAP Theorem.

- SQL, NoSQL, and NewSQL systems.

- Database as a service (DaaS).

Topics

Data Processing:[2]

- Batch processing (e.g., Apache Spark).

- Stream processing (e.g., Apache Flink).

- Industrial-scale machine learning (e.g., Apache SystemDS).

Emerging Trends:[2]

- Self-designed and learned data-intensive systems.

- Blockchains and database management systems.

- Data management on modern hardware.

---

[2]Topics not fixed yet.

Topics

We will cover some basics of database management systems, but **not the internals in detail.** For internals, we refer to other courses at the Department of Computer Science.

Undergraduate Courses:

- Databases 1 (VO+PS; summer semester)
- Databases 2 (VO+PS; winter semester)
- Database Tuning (VO+PS; summer semester)

Graduate Courses:

- Advanced Databases (VO+PS; winter semester)
- Non-Standard Database Systems (VO+PS; summer semester)
- Similarity Search in Large Databases (VO+PS; winter semester)

Course Objectives

- Develop an **intuition** for database systems and their **core principles.**
- Be able to **properly use the terminology** that is commonly used in these systems.
- **Tackle simple problems** using these systems.
- Learn to **choose a system** for a given problem/application context.
- **Hands-on experience** with selected systems.
- Working in a **team** (organize yourself as a team).
- Learn about the **challenges and limitations** of these systems.

Prerequisites

VU "Grundlagen Informatik und Systeme" (512.023)
  ⇒ **Core principles** of computer science and systems.

UE "Einführung in Programmieren mit Python" (512.024)
  ⇒ **Basic programming skills** (e.g., in Python3).

Logistics

Website: https://dbresearch.uni-salzburg.at/teaching/2023ss/dim/

- Announcements (important announcements also via email).
- Detailed grading scheme.
- Learning material: slides, hands-on instructions, and supplementary material.
- Assignments: schedule, meetings, and late submission policy.
- Midterms: schedule, topics, and modes.

Where to Find Help

Slack:

1. **Create an account**[3] for our DBTeaching workspace (using your stud email).

2. **Log in**to the workspace.

3. Browse the channels and **search** for **channel** "dim-uv-2023ss".

4. **Join channel** "dim-uv-2023ss".

5. **Say "Hi"** to the others in the channel 😄.

---

[3]Top right corner in most browsers.

Lecture

General Information:

- There are **two groups:**
    - Group 1: **Mondays, 10:00 - 12:00 pm CET**, Hörsaal I – Christian Doppler.
    - Group 2: **Wednesdays, 01:00 - 02:30 pm CET**, Hörsaal 2 – Lise Meitner.
- Lecture will be in **hybrid mode**, i.e., **in person** but **streamed online.**
- However, **attendance** is **mandatory** and will be checked – *"prüfungsimmanent"*.
- **Recordings** will be available via Google Drive.
- **Theoretical background** for the **assignments.**
- Covers all **relevant topics** for the **exams.**

Assignments

General Information:

- **Groups of three** students (enrollment via Blackboard[4]).
- **Practical assignments** related to the topics covered in class.
- Submission via **Blackboard** (5 weeks per assignment; max. 6 weeks).
- **3 assignments**, **each of which contributes 18%** to your grade.
- Initially graded per group, **individual grading** may apply after meeting.
- Please notify me if a student does not contribute to the assignments.

---

[4]If there is any problem in the Blackboard course, please notify me as soon as possible.

Assignments

Late Submission Policy: You can submit late (up to one week) but **you will lose 7%** of the assignment's total points for **every 24h delay**. Delay is computed with respect to the initial deadline and is **rounded up** to the next multiple of 24.

Example: Assignment 1 has max. 18 points and is submitted 25h late. Then, the delay is rounded up to 48h and the max. points are $18 \cdot (1.0 - 0.07 \cdot 2) = 18 \cdot 0.86 = \underline{15.5}$.

After-Assignment Meetings: Short meetings to **discuss your submission** and the **grading**. **One meeting per assignment and group** (max. 15 minutes). Students will need to **answer questions** directly related to the assignment. This may also affect the final individual grading (positively or negatively).

Assignment 0 – Get Familiar with the Workflow

Tasks:

1. Find yourself a **team of three students.**
2. **Enroll your team** in the Blackboard course.
3. Play around with **Debian Linux** (optional).
4. **Submit the answers** to the questionnaire.

Deadline: March 22, 2023, 11:55 pm (aka 23:55) CET

Meeting in CW 13 and **no late submission** for this assignment.

Students without a team are randomly assigned to groups.

General Information:

- Exams will be held via **Blackboard**.
- **2 exams** (midterm and final), **each of which** contributes **23%** to your grade.
- An exam will last **at most 90 minutes.**

The **exam date** and **time** is the **same for all students** (disregarding PlusOnline groups).

Schedule Overview

CW	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27
		March				April				May				June			July	

	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27
Ass.	Assignment 0			M			Assignment 2					M					M	
			Assignment 1					M			Assignment 3						M	
Exams									QA	MT							QA	FE

M   After-Assignment Meetings (and Late Submission, Q&A for the subsequent assignment)

MT   Midterm

FE   Final Exam

QA   Question & Answer session

CW   Calender Week

Grading Scheme

Assignments		Exams		Total
Assignment 0	0%			
Assignment 1	18%			
Assignment 2	18%	Midterm exam	23%	
Assignment 3	18%	Final exam	23%	
	54%		**46%**	**100%**

Grading Scheme

Overall Points	Grade
≥ 88.75%	1 – "Sehr Gut"
[77.5%, 88.75%)	2 – "Gut"
[66.25%, 77.5%)	3 – "Befriedigend"
[55%, 66.25%)	4 – "Genügend"
<55%	5 – "Nicht Genügend"

Q&A

Computer Science Terminology

In-class Exercise:

1. Find yourself a group or work alone – **as you prefer.**

2. The next slide lists **28 + 1 terms.**

3. Try to briefly **describe each term** in the **context of computer science.**

Time: 30 minutes

Afterwards, we discuss these **terms in the plenum.**

Operation(s)	Instruction(s)	Time
Concurrency	Parallelism	Space
Processor / CPU	Main memory (RAM)	Cache
Efficiency	Performance	Transparency
Trade-off	Virtual machine	Redundancy
HDD	SSD	Memory hierarchy
Database system	Query	Index
Transaction	Scalability	Physical level
Declarative	Imperative	Compiler
Address		