UV Distributed Information Management

Introduction

Daniel Kocher

Salzburg, March 6, 2024

Department of Computer Science
University of Salzburg

database research group

PARIS LODRON UNIVERSITÄT SALZBURG

Agenda

Today's Agenda

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

Personal Information, Contact & Lecture Policies

Who Am I? **Daniel Kocher** (Research Assistant, Comp. Systems / Database Research Group)

Contact (in this order):

1. **Lecture:** Wednesdays, 01:00 - 02:30 pm CET.
2. **Slack:** https://dbteaching.slack.com/ (**create** an account with your 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 continuously 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.

Topic Outline

Distributed Information Management

Data Management

Data Processing

Distributed Information Management

Data Management

Data Processing

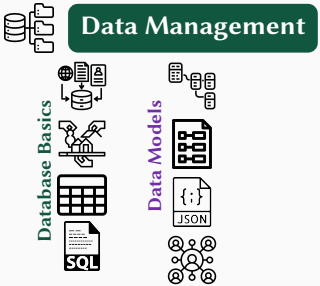
Purpose

Distributed Information Management

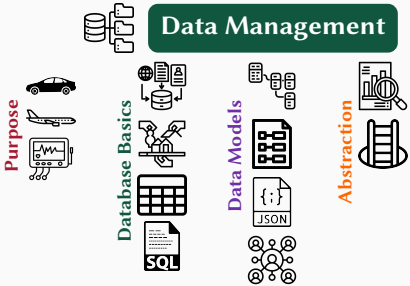
Data Management

Data Processing

Distributed Information Management

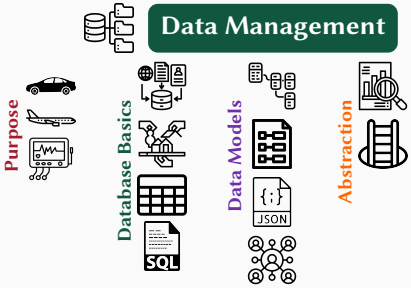


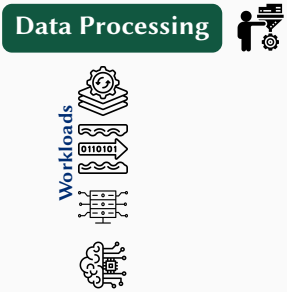
Distributed Information Management



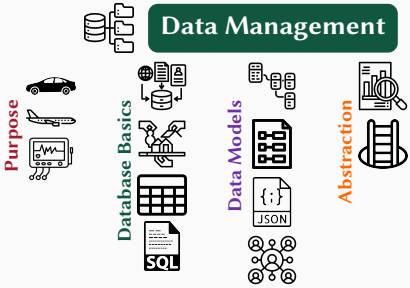
Data Processing

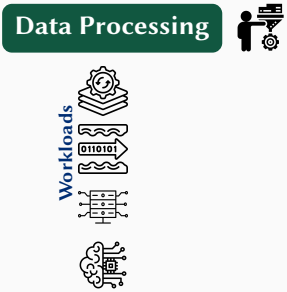
Distributed Information Management



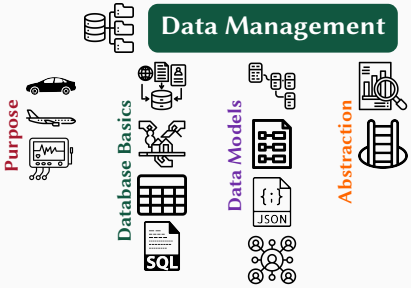


Distributed Information Management





Distributed Information Management



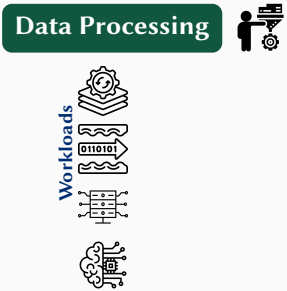


Figure 1: Generated using Stable Diffusion WebUI:
https://github.com/AUTOMATIC1111/stable-diffusion-webui

Complementary Courses

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

Where to Find Information

Website[2]:

- Assignments: schedule, meetings, and late submission policy.
- Detailed grading scheme.
- Learning material: slides, hands-on instructions, and supplementary material.
- Midterms: schedule, topics, and modes.

Blackboard[3]:

- Announcements (important announcements also via email).
- Assignments: announcement and submission.

[2] Course website: https://dbresearch.uni-salzburg.at/teaching/2024ss/dim/
[3] PLUS Blackboard: https://elearn.sbg.ac.at

Where to Find Help

Slack:

1. **Create an account**[4] for our DBTeaching[5] workspace (using your stud email).
2. **Log in**to the workspace.
3. Browse the channels and **search** for **channel** "`dim-uv-2024ss`".
4. **Join channel** "`dim-uv-2024ss`".
5. **Say "Hi"** to the others in the channel 😄.

[4] Top right corner in most browsers.

[5] DBTeaching Slack Workspace: https://dbteaching.slack.com/

Lecture

General Information:

- Only **one group: 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 **obligatory** 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 2-3** students (enrollment via Blackboard[6]).
- **Practical assignments** related to the topics covered in class.
- Submission via **Blackboard** (4 weeks per assignment; max. 5 weeks).
- **3 assignments**, **each contributing 20%** 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.

[6] 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. 20 points and is submitted 25h late. Then, the delay is rounded up to 48h and the max. points are $20 \cdot (1.0 - 0.07 \cdot 2) = 20 \cdot 0.86 =$ 17.20 points.

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). **Every student** must be able to **answer questions without help!**

Assignment 0 – Get Familiar with the Assignment Workflow

Tasks:

1. Find yourself a **team of 2-3 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 13, 2024, 11:55 pm (aka 23:55) CET

Optional meeting and **no late submission** for this test assignment.

Students without a team after the deadline are **randomly assigned to groups.**

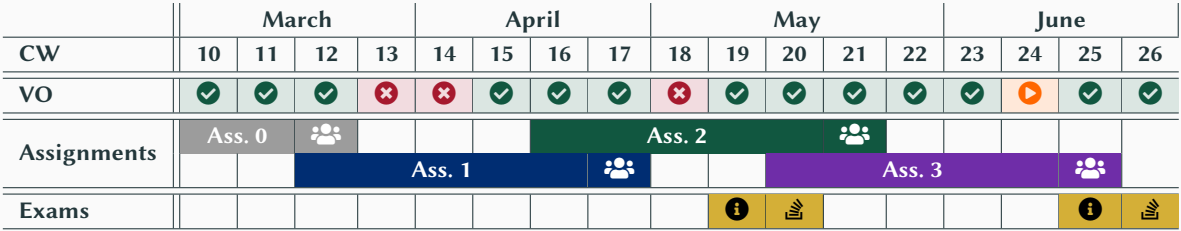
Exams

General Information:

- Exams are **open book** and will be held **in person** (justified exceptions may apply).
- **2 exams** (midterm and final), **each of which** contributes **20%** to your grade.
- An exam will last **at most 90 minutes.**

The **exam date** and **time** is the **same for all students**.

Schedule Overview



CW Calender Week

- Regular lecture (in person / hybrid)
- No lecture
- Remote lecture or video recordings to watch
- After-Assignment Meetings
- Midterm / Final Exam
- Question & Answer session

Grading Scheme

Assignments		Exams		Total
Assignment 0	0%			
Assignment 1	20%			
Assignment 2	20%	Midterm exam	20%	
Assignment 3	20%	Final exam	20%	
	60%		**40%**	**100%**

Grading Scheme

Overall Points	Grade
$\geq 88.75\%$	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.**

Throughput
Concurrency
Processor / CPU
Efficiency
Trade-off
HDD
Database system
Transaction
Declarative
Address

Instruction(s)
Parallelism
Main memory (RAM)
Performance
Virtual machine
SSD
Query
Scalability
Terminal

Time
Space
Cache
Transparency
Redundancy
Memory hierarchy
Index
Physical level
Compiler

Credits

Icons (pictograms made by)

mynamepong	Freepik	juicy_fish	Uniconlabs
Smashicons	Eucalyp	Elzicon	Skyclick
Muhammad Atif	surang	ultimatearm	imaginationlol
Pixel perfect			

from https://www.flaticon.com