Databases

BCS1510

Dr. Katharina Schneider &

Dr. Tony Garnock-Jones*

📅 Week 1 - Lecture 1

📍 EPD150 MSM Conference Hall

Adapted with gratitude from the original lectures by Dr. Ashish Sai

*CC both of us when emailing please!

About me

Academic Background
- BA: Dual Studies (FH Aachen)
  - Studies: Scientific Programming
  - Training: Software Developer
- MA: Operations Research (DACS at Maastricht University)
- PhD Thesis: On Agent-Based Modeling of Sex, Plants and Sustainability

Now
- Lecturer: Projects, Computer Science Skills, Databases, Operations Research (Aachen)
- Research: Cancer-Modelling, Agent-based Models, Educational Research

I love
- My Family
- Soccer (yes, Bayern München is the best team in the world)
- Chocolate (in fact, I am a chocoholic)
- Being outside when sun is shining

Maastricht University

Changes to the course this year

- Smaller, better structured slides with less pages ✅

- Practice "mid-term" exam, more practice exams ✅
  - We will give you a mid-term exam (not graded) to help you estimate your learning success
  - Multiple practice final exams

- No big group project ✅

- "Do not ask us to ask TA for help" – no, you still have to ask the TAs
  - Not possible for us to answer all the questions in the lab on our own (or let Ashish do that) (the reason we have TAs)

Changes to the course this year

- 5 Minutes break is too short ✅
    - Fine, you can have 6 minutes…! Okay, let's say 10.

- Clearly tell us what is important for exam ✅
    - Everything covered in the lectures, tutorials and labs (Sorry 😔)
    - We will give you a list of intended learning objectives before each lecture.

- (More) Transparent Grading Scheme for Project ✅
    - Project specification now contains a detailed grading rubric

- Talk a bit slower ✅
    - Tony says he often struggles with this so please remind him when it becomes a problem

- Better Jokes (I am German, do not expect anything here) 😓

Maastricht University

First Lecture ☺

Learning Objectives

By the end of this lecture, you should be able to:

1. Understand the Role of Databases:
   - Explain the need for databases in organizing and managing data.
2. Apply System Design Principles:
   - Utilize concepts from previous courses (OOM) to design and evaluate information systems that meet specific operational requirements.
3. Grasp Database Fundamentals:
   - Define key components of database systems including databases and database management systems (DBMS).
   - Identify different types of DBMS and understand their applications across various industries.
4. Understand Database Operations:
   - Understand essential database operations such as data insertion, deletion, and querying within the context of an online bookseller.
   - Recognize challenges in data management such as ensuring data consistency, atomicity, isolation, and durability.
5. Course Structure:
   - Understand the course objectives and structure

Imagine: I am a mafia boss [1]

[1] Purely hypothetical!

What's a boss gotta have in his toolbox, eh?

- I want to organise my group of "picciotti"

- To achieve more efficiency in all our operations

Efficient Mafia 🕵️

- Despite what you might think

- Being Mafia boss requires a lot:
  - Security
  - Privacy
  - Book-keeping (might be risky, but hey, some like control)

Maastricht University

Efficient Mafia Data: Personnel & Partners

- I need to store information about

  - People that work for me (Soldiers, caporegime, etc.) 👥
  - Organizations I do business with (police, 'Ndrangheta, politicians) 🤝
  - Assignments of soldiers to operations

Efficient Mafia Data: Operations & Guidelines

- I need to store completed and open operations
  - Protection rackets 🛡️
  - Loan sharking 💰
  - Control of contracting/politics 🏗️

- I need to avoid any of my people being involved in:
  - Burglary, mugging, kidnapping (too much police attention) 🚓

- Cover-up operations/businesses 🕵️

Maastricht University

Efficient Mafia Data

- Data Sharing with External Organizations, protection some of the information 🤝
    - For business purposes

- Thus, I need Access Control 🔑
    - Boss, Underboss, and Consigliere (advisors) have full access and need to be able to:
        - Assign soldiers to operations.
        - Manage operations (create/shutdown)
        - Financial transactions (e.g., paying cops)
        - Monitor financial flows

Efficient Mafia Data

- I need Accountant Access 💼 (all 20 of the accountants)
  - Money bookkeeping
  - Manage money laundering operations
  - Bank transfers
  - Record bribing expenses

- Daily Reporting by Soldiers (6424) 📊
  - Misdeeds in a daily log 📝
  - Money: expenses and collections 💰

Maastricht University

Efficient Mafia Data

- Semi-Public Interface for other bosses I collaborate with 🤝

  - Search for cops on our books 👮
  - Check controlled areas 🗺️
  - Plan coups d'état 🔄

Maastricht University

The offer…

You've been handpicked by the family, kid.
Your job? To architect our information system. Do us proud.

Mafia Information System 💼

- Success: 💵 + 🚗 + 🎉 = Great Life!

- Failure: 😨 Let's not go there...

Maastricht University

How can you design this information system?

- Recall modelling and designing from your OOM course

- What to represent (key entities)?
    - Attributes and Operations

Maastricht University

How can you design this information system?

- How to store the data? 📁
  - Maybe, we can just use files: people.txt, organisations.txt, operations.txt, money.txt, daily-log.txt

- Access Control 🔐
  - Fine-grained permissions.
  - Accountants see `money.txt` only (no names and addresses of our soldiers) to prevent… ambitious ideas (takeovers) 💡 ➡️ 🚫

More problems for you – Data Access

- How to access data? :
  - Write a program that opens one or more files, scanning through them and reading/writing information in them 📚

- Access patterns and performance
  - How to identify shops overdue on payments for over a month. ⏳ 💰
  - Scan huge file, sort by time, pick the oldest, measure time?

Maastricht University

More problems for you

- Atomicity 💵
  - Ensure money transfers complete fully or not at all
  - *"Money vanishes only if you plan to disappear too."*

- Consistency 📊
  - Maintain data in a valid state at all times.
  - E.g. no two operations share a name.

Maastricht University

More problems for you

- Isolation 🚫 🔄
    - Soldiers add to `daily-log.txt` simultaneously.
    - Prevent overwrites to avoid ~~firing~~ for low productivity

- Durability 💾 ✅
    - In case of a computer crash, ensure data persists and remains accurate after crashes
    - Payments to cops must be reliably recorded (avoid double payments).

Maastricht University

Summary: What are the requirements for our Mafia information system?

- Data Access
- Access Control
- Atomicity
- Consistency
- Isolation
- Durability

Would a file-based approach work?

- What do we mean with File-Based System 📁
  - A collection of application programs that perform services for the end-users, such as the production of reports. Each program defines and manages its own data (Connelly and Begg)

  - Early attempt to computerize the manual filing system that we are familiar with
    - In your daily life: receipts, warranties, invoices, bank statements...

Maastricht University

Limitations of a file-based system

- Separation and Isolation of Data
  - Data is stored in different files
  - Makes cross-referencing or combining data challenging
- Duplication of Data
  - Wasteful, costs time and money
  - Leads to loss of data integrity (inconsistent data)
- Data Dependence
  - Tight connection between data and the programs that process it
  - Makes changes to the files structure or format challenging
- Incompatible file formats
  - Structure of files is embedded in the application programs -> structures are dependent on application programming language
  - Creates the need of converting files to common format (time-consuming)
- And many others

THE SOLUTION

- Thankfully, there is something called Database Systems. These are designed to overcome these limitations

Introduction to Database Systems
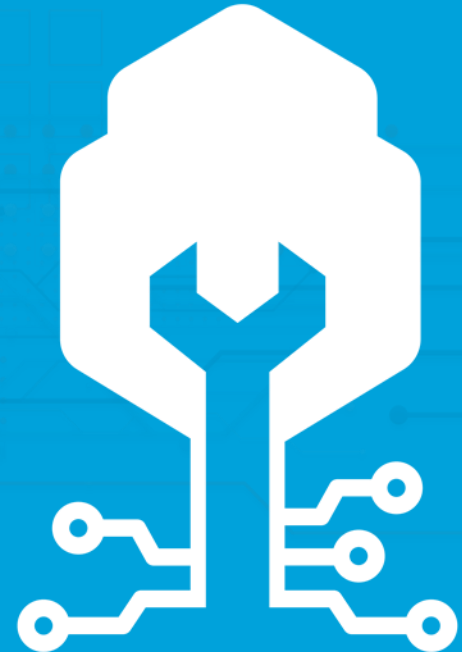
Database Systems

- Database
  - A collection of related data

- Database Management System
  - The software that manages and controls access to the database

- Database Application
  - A program that interacts with the database at some point in its execution

Maastricht University

Where do we use Database Systems?

- Any thoughts?

Where do we use Database Systems?

- Buying groceries from Supermarket
  - Product Inventory
  - Point of Sale System
  - Customer/rewards database
  - …
- Spotify and Netflix Subscription
  - User accounts & authentication database
  - Content Catalog Database
  - Recommendation system (ML database)
  - …
- Buying stuff using your credit card
  - Check whether you have enough credit left to make the buying
  - Credit card on the list of stolen cards?
  - …
- Booking a vacation with travel agent
- Using the local library
- More can be found in Connolly and Begg

Database Definition (1)

Ullman and Widom:

- Term database refers to a collection of data that
    - Is stored over a long period of time
    - and managed by a data base management system (DBMS)

Database Definition (2)

Connelly and Begg:

- A shared collection of logically related data and its description, designed to meet the informational needs of an organization.

Database – Examining the Definition

- Single repository of data that can be used simultaneously by many users
- Minimum duplication due to integration of all data items
- Self-describing nature of a database (data + metadata)
- Data abstraction and separation of data structure from the application (internal definition of an object + separate external definition)
- Information takes the form of *entities*, *attributes* and *relationships*.

What do you mean by Logically Related?

- When identifying Information Needs, we attempt to identify:
  - Entities 🧩
    - Distinct object in an organization (person, place, …) 🏢 👤
  - Attributes 📝
    - Property describing aspects of an entity (age, height,…) 📊
  - Relationships 🔗
    - Association between entities (has, buys, sells, returns,…) 💼 ⚖️

- Entity relationship diagrams illustrate these concepts (More in this/next week's lab) 📈

Entity Relationship Diagram for Mafia
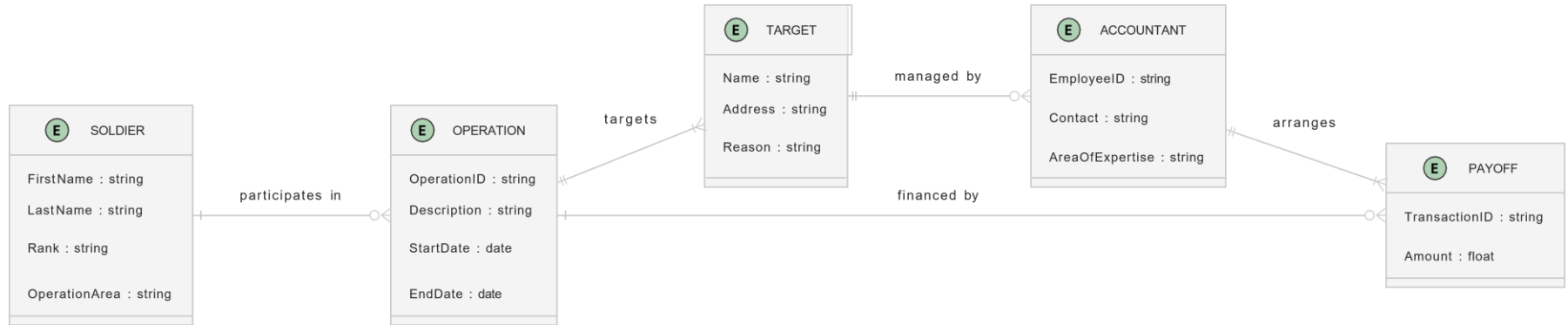
More Mafia and I might get in trouble

- More class friendly example: Amazon!
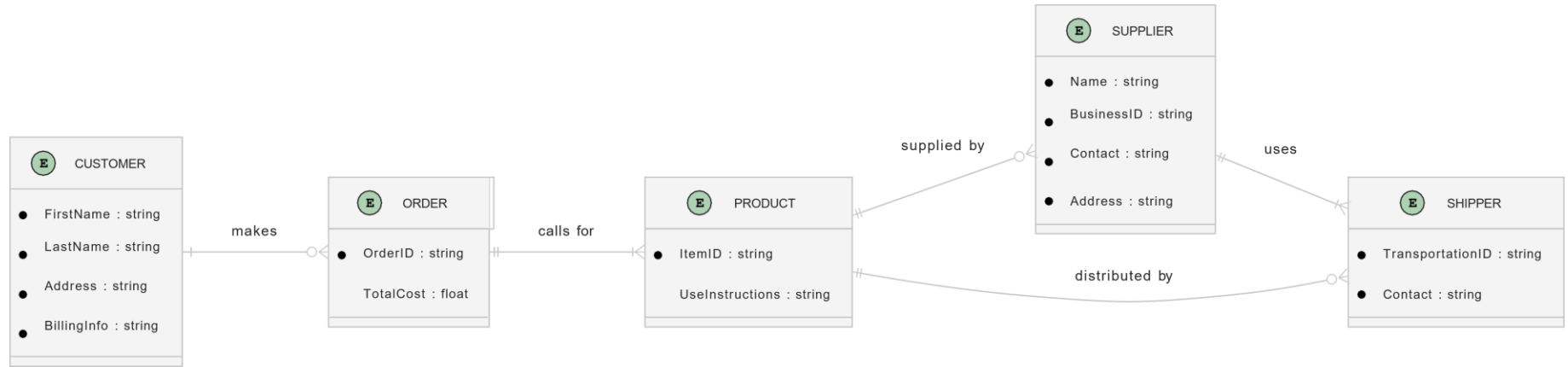
Any doubts about what database is and what we mean by logically related?

Short Break

- **10** minutes

Remember: Database Systems

- Database
  - A collection of related data

- Database Management System
  - The software that manages and controls access to the database

- Database Application
  - A program that interacts with the database at some point in its execution

What is a DBMS?

- Definition (Connelly and Begg)
  - A software system that enables users to define, create, maintain and control access to the database

What is a DBMS

A DBMS is expected to

- 📝 Allow users to define the logical structure of a database through a data-definition language (DDL)

- 🔄 Give users the ability to insert, update, delete, and retrieve data from the database through a data-manipulation language (DML)

What is a DBMS

A DBMS is expected to

- 💾 Support the storage of very large amounts of data over a long period of time, allowing efficient access to the data for queries and database modifications
- 🛡️ Enable durability, the recovery of the database in the face of failures, errors of many kinds, or intentional misuse
- 🤝 Control shared access to data from many users at once

Examples of DBMS

- Oracle, MySQL, IBM DB2 (Relational DBMSs)
- SAP Hana, Microsoft SQL Server (Relational DBMSs with in-Memory Capabilities)
- Microsoft Access, PostgreSQL, SQLite (General Purpose Databases)
- Apache Hbase, MongoDB, Neo4j (NoSQL Databases)

- We will focus on relational DBMSs

Remember: Database Systems

- Database
  - A collection of related data

- Database Management System
  - The software that manages and controls access to the database

- Database Application
  - A program that interacts with the database at some point in its execution

Maastricht University

(Database) Application Programs

- A computer program that interacts with the database by issuing an appropriate request (typically an SQL statement) to the DBMS

Database System

Database Application

Database Application

DBMS

Database

Database

Maastricht University
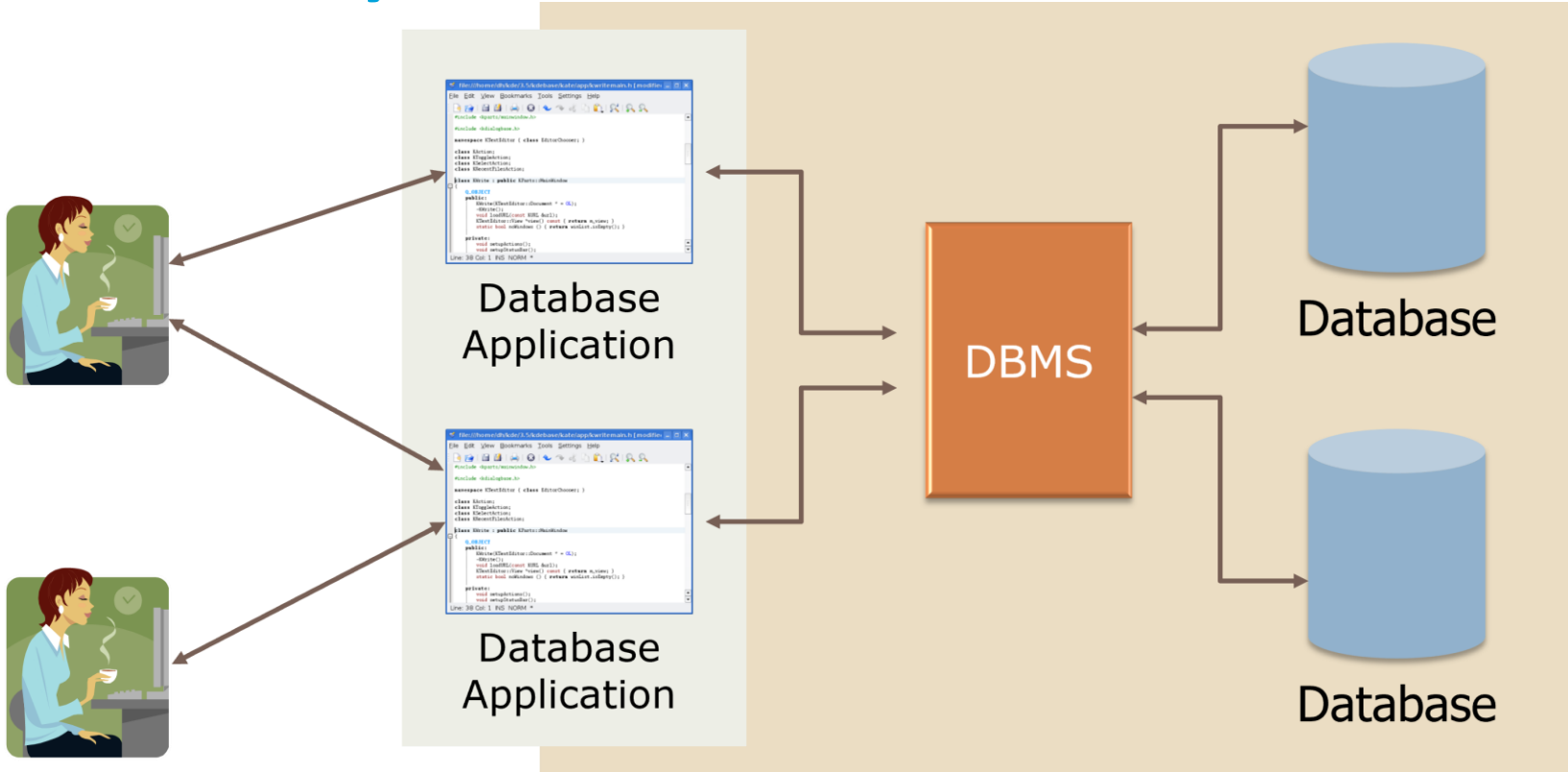
Recap

- What is the difference between a database and a DBMS?

Maastricht University

Recap

- Database: Collection of logically related data
- DBMS: Helps in defining and maintaining this collection

Maastricht University

Example: Online Bookseller

- Business Model 📚
  - Buy books, store them, and sell via a website.
- Essential Data 💾
  1. Books 📖
  2. Customers 👥
  3. Orders 🛒
  4. Preferences ⭐

Maastricht University

Example: Online Bookseller

- Data operations
  - Insert ➕
  - Remove ➖
  - Search by author/title 🔍
  - Change status (stock) 🔄

Can you imagine data related issues?

- Data Accuracy and Consistency
  - E.g. if inventory records aren't updated in real-time, a book might be listed as available when it's actually out of stock.
- Data Redundancy
  - E.g. a book may have multiple records in the database if it's listed by different authors, publishers, or categories.
- Data Integrity
  - E.g. if two people try to update the stock quantity of the same book at the same time

What does the DBMS do in the Bookseller Example?

- 📚 Describe real-world entities in the database (e.g., Books) using the Data Definition Language (DDL)

- ✍️ Read and Write to this database (e.g. add a sale and reduce the stock) using the Data Manipulation Language (DML)

- 🗃️ Stores data over a long period of time (users may ask for a years-old receipt)

What does the DBMS do in the Bookseller Example?

- 💪 Ensures Durability (What if a user places an order and is paying for the book but our website stops due to a crash? What happens to the order in the database?)

- ✍️ Ensures Concurrency (What if four users are trying to buy Darth Vader Autobiography at the same time and we only have one copy?)

Conclusion

- Databases/DBMSs are useful, right? RIGHT?
- Let's have a look at some history

Rise of Relational Database Systems

- Early DBMSs
  - were mostly designed for hierarchical or network data models, which had limited querying capabilities compared to modern relational databases

- Relational database systems
  - Edgar Frank "Ted" Codd, 1969: devised the relational model
  - Queries could be expressed in a very high-level language
  - Norm by 1990

Maastricht University

Database Technology Evolution Timeline

Database Technology Evolution Timeline
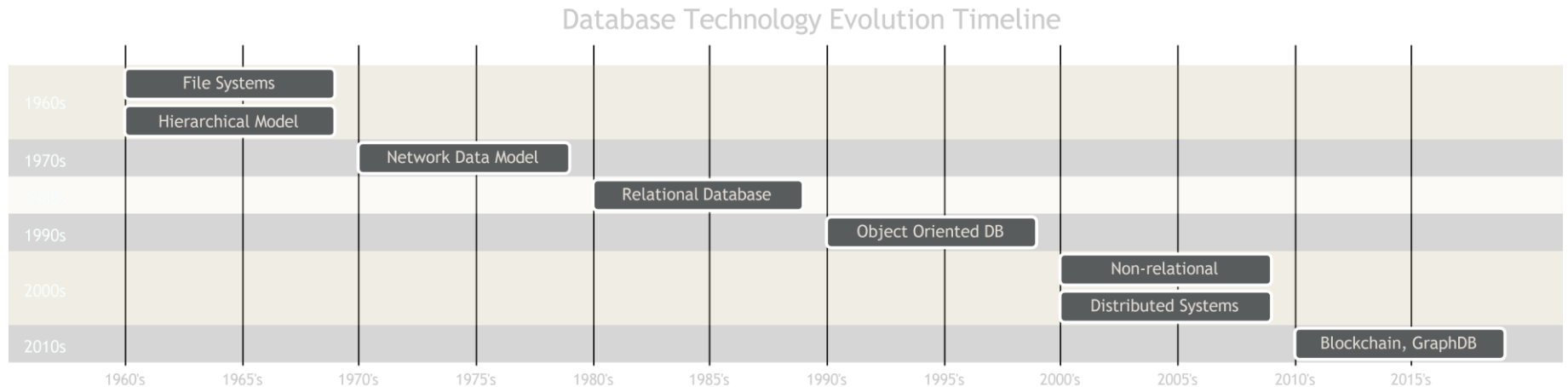
Why this Database course?

To help you run an efficient mafia empire? 🕵️

Absolutely **NOT** [2]

[2]: No guarantees though!

Why this course?

- Moore's Law

- Is anyone aware of what Moore's Law is?[3]

- [3]: I hope the answer is a clear yes.

Why this course?

- Moore's Law: Computers double their processing power roughly every 18 months, and it might be even quicker now. 💻 ⚡

- Data Explosion: The amount of data we create doubles every nine months. 📈

- The Need for DB's: With data growing faster than we can keep up, we rely on computers to help us understand and manage it. 🤖 📊

Why this course?

- Database is no longer just about maintaining (boring) library and payroll systems. 🚀

- Today, databases are the powerhouse behind virtually everything you interact with online:
  - 🎥 Netflix   -
  - 🛒 Amazon's Product Catalog
  - 🔍 Google Search
  - 🎮 Fortnite

Why to study for that course

- DACS: it is mandatory
- Why databases matter:
  - At the heart of many technologies:
    - 🌐 Web Search
    - 🔍 Data Mining
    - 📊 Big Data
    - 🔗 Blockchains
    - 🕵️ And possibly… the Mafia?

Convinced?

- No?????

- Let's get serious then

Maastricht University

1. Data analyst [↗]

National average salary: $70,676 per year [↗]

Primary duties: A data analyst [↗] collects, cleans, analyzes and visualizes data to derive insights. They use tools such as spreadsheets, SQL and Tableau to identify trends and patterns in data and use the results to help organizations make data-driven business decisions. They also create reports and visualizations to communicate their findings to stakeholders. Data analysts have strong analytical and problem-solving skills and communicate complex information to non-technical audiences. These specialists can work across various industries, such as health care, finance, marketing and retail.

5. Data modeler [↗]

National average salary: $102,376 per year [↗]

Primary duties: A data modeler [↗] designs and maintains the data architecture of an organization, which involves creating and managing complex data models that facilitate the storage, manipulation and retrieval of data. The duties of a data modeler can include analyzing business requirements, defining data architecture standards, creating data models, developing data dictionaries and ensuring data integrity and security.

6. Data scientist [↗]

National average salary: $124,693 per year [↗]

Primary duties: A data scientist analyzes, interprets and organizes large and complex data sets. This involves gathering, cleaning and transforming data into a form others can analyze using various statistical techniques and machine learning algorithms. Data scientists have excellent programming skills, proficiency in statistical modeling and data analysis and a solid understanding of database technologies. They also work with stakeholders to understand business needs and provide insights that help inform business decisions.

7. Data architect [↗]

National average salary: $125,977 per year [↗]

Primary duties: A data architect [↗] designs and implements data solutions for organizations. They analyze business requirements, design and implement data models, define data governance policies and ensure data security and privacy. Data architects also oversee the implementation of data solutions, monitor performance and ensure the data architecture remains up-to-date and relevant. Additionally, they often collaborate with cross-functional teams and manage data-related projects from conception to completion. They also work closely with stakeholders to understand their data needs and create solutions that meet those needs while aligning with the organization's overall business strategy.

8. Big data engineer [↗]

National average salary: $128,631 per year [↗]

Primary duties: Big data engineers manage and analyze large and intricate data sets and design and maintain data pipelines to transfer data from diverse sources into the company's data storage systems. Additionally, big data engineers have expertise in various data management technologies, such as Hadoop, Spark and NoSQL databases.

What will you learn in this course?

- We will scratch years of research on databases in a few months.

- You will be able to understand how relational database management systems function.

- You will also be able to use SQL.

What will you learn in this course?

Lecture	Title	Attention Level
Lecture 1	Introduction	Easy stuff 😊
Lecture 2	The Relational Model and Basic SQL	Still easy 😌
Lecture 3	Data Manipulation with SQL (Part 1)	Pay Attention 👀
Lecture 4	Data Manipulation with SQL (Part 2)	Pay Attention 👀
Lecture 5	Data Manipulation with SQL (Part 3)	Pay Attention 👀
Lecture 6	SQL Constraint	Pay Attention 👀
Lecture 7	Design Theory of Relation Databases (Part 1)	Pay some serious attention 🧐
Lecture 8	Design Theory of Relation Databases (Part 2)	Pay some serious attention 🧐
Lecture 9	Query Tuning	Pay attention 👀
Lecture 10 and 11	Beyond the Relational Model	-

Course structure

- Classes: twice a week, Tue + Wed either 08:30 or 11:00
  - Look *closely* at your timetable
  - The lectures are **not** at the same time each week
- Labs: once a week, Fri 08:30
  - Used to work on the project
- Tutorials/Assignments
  - To be done at home
  - Questions? Ask on Discord or in lecture
- Project
  - Group work in team of 2

Course Material

- Textbooks (no need to buy the book)
  - J. D. Ullman and J. Widom, A First Course in Database Systems, 4th edition, Pearson, 2014
  - Thomas Connolly and Carolyn Begg, Database Systems, 6th ed., Pearson, 2014

- Online Resources
  - All course material is available on Canvas
  - Discord will be actively monitored (Q&A)

Lab Resources

- MySQL with DBeaver
  - Set up guide: Lab 1

Bugs/Typos

- Please report any suspect bugs/typos
    - Lecture slides 1-5: Katharina
    - Lecture slides 6-11: Tony
    - Labs: Tony
    - Project: Tony

- We need your help to improve the course material!!

Time Investment – What is expected from you

Category	Number	Time Each	Total time
Class time			
Lectures	11	2	22
Labs	6	2	12
Revision class	1	2	2
Total in Class Time:			36
Outside Class Time			
Practice Exams	1	2	2
Going through lecture slides/notes	11	2	22
Revision			30
Project Work	1	22	22
Total Outside Class Time			76
Total Time			112

Assessment Components

- Final exam: 75%

- Project: 25% (Group of 2) (if you pass the exam)
  - Deadline: May 12, 2025
  - 1 point penalty for late submissions for each day
  - No submissions will be accepted after the final exam
  - Rubric available on Canvas

Maastricht University

Project (Team up!)

Cheapo Software Solutions (CSS) wants to monetise Quackstagram by selling user data analytics

You have to create a Database for Quackstagram and help CSS answer questions about their users (their preferences to sell them products)

Maastricht University

Grading scheme

Grade	Range
10	>= 95% - <= 100%
9	>= 85% - < 95%
8	>= 75% - < 85%
7	>= 65% - < 75%
6	>= 55% - < 65%
F	< 55%

You have to pass the exam to get the grades for the projects

You need >=5.5 in your exam (more than 55% in the written exam)

Academic Integrity

- Anything you submit for credit is required to be your own work.

- It is okay to share ideas but not the solution.

- Claiming others work is a misbehaviour.

- We implicitly trust you, but a breach will be reported to the university.

ChatGPT and other LLMs

- Fine to generate test data etc.

- You will not learn if you make LLMs "think" for you

  - This is easy to do by accident even if you think you're not ☹

- If you use LLM you **_MUST_** disclose the usage

  - You are committing Academic Dishonesty if you do not (look it up!)

- Do not use LLMs to write code/SQL (it will not help you for your exam)

Maastricht University

Don't be like this
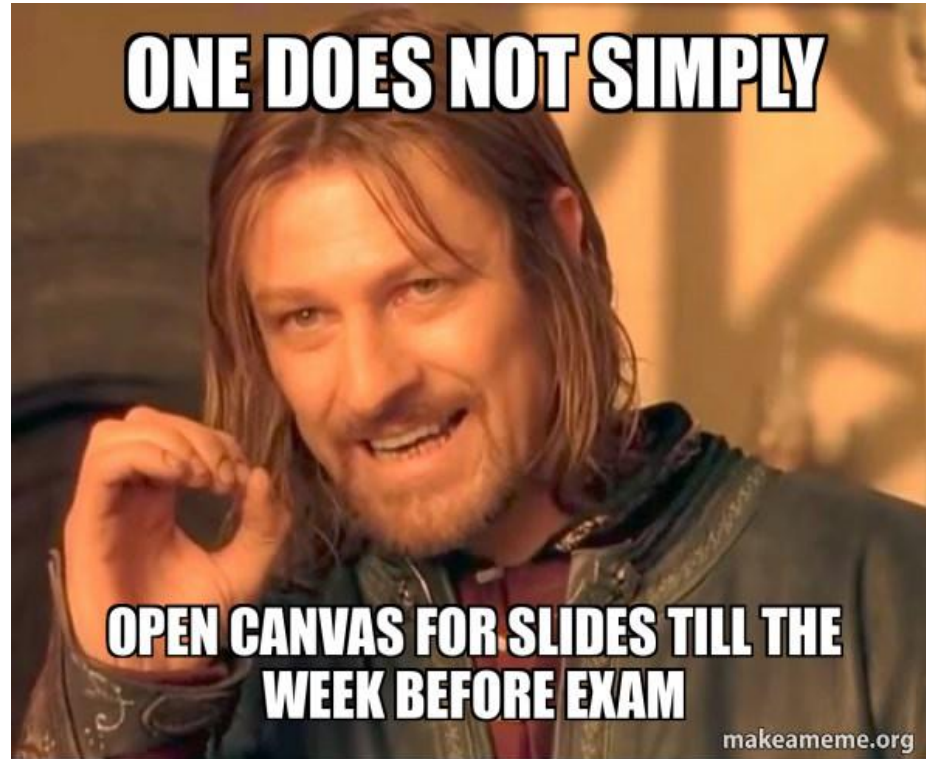
Do this!!!

Contact details

- If you have any issues throughout the course, contact us via email:
- Tony: tony.garnock-jones@maastrichtuniversity.nl
- Katharina: k.schneider@maastrichtuniversity.nl
- Or send a message on Discord
- (https://discord.gg/Be2KSF8QG6)

Questions?

See you tomorrow