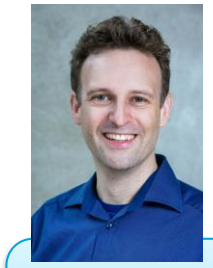
Web and Database Technology

For now: Into Databases…!

Christoph Lofi
Ujwal Gadiraju

TU Delft WIS Web Information Systems

Christoph Lofi



Albert Power

CSE1500: Database Technology

50%

Midterm Exam

Assign. DB.1

Assign. DB.2

Assign. DB.3

Ujwal Gadiraju

Albert Power

CSE1500: Web Technology

50%

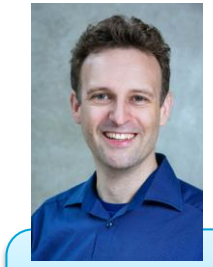
Endterm Exam

Assign. Web.1

Assign. Web.2

Assign. Web.3

Asterios Katsifodimos



CSE1505 Information & Data Management

100%

Endterm Exam

Assign. 1

Assign. 2

Assign. 3

Web and Database Technologies

- Grading
  - (Database) **Midterm** 50%
  - (Web) **Endterm** 50%
  - **Assignments**
    - See Brightspace
    - Assignments are not graded
    - Assignments will help you practice the more theoretical skills and teach you hands-on practical skills.
      - They are VERY VERY important.
      - Research shows that the failing the course is strongly correlated to not doing the assignments (properly)…

Exams

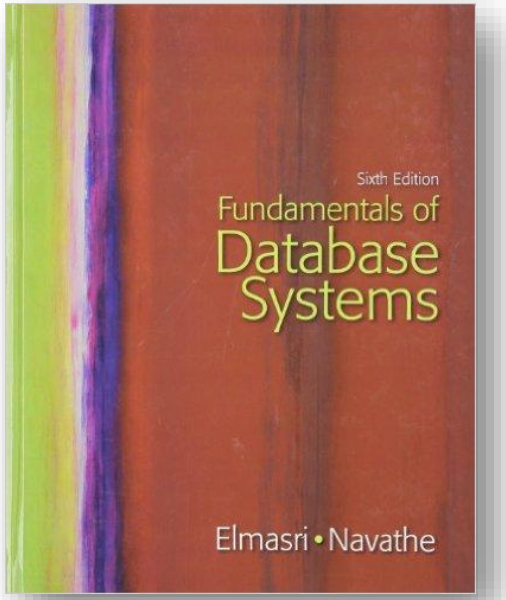
- **Midterm** Exam focuses on DB
- **Endterm** Exam focuses on Web
- Each exam can be Resit
  - either both or individually
  - New faculity rule starting 25/26:
    - To be eligible for cum laude, you can only resit a passed exam in the same academic year in which you first passed it.

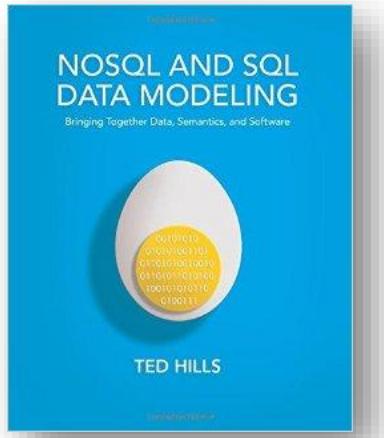
More Information

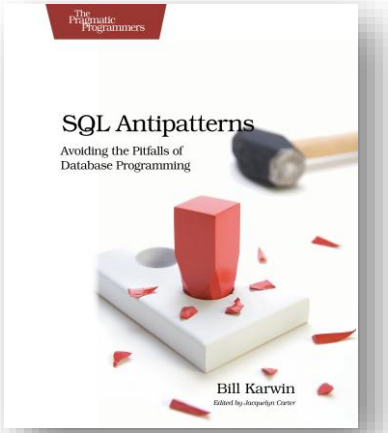
- Prerequisites
  - Basic understanding of OOP languages
  - **YOU NEED TO REGISTER ON BRIGHTSPACE!**
- Materials
  - Slides (on Brightspace)
  - Video Recordings (on Brightspace)
  - Extra Materials (on Brightspace)
  - Lab assignments
  - Discussions on https://answers.ewi.tudelft.nl/categories/1/tags/152
    Use at least tags “CS1500” with “2526”
- Syllabus on Brightspace
  - Should cover most questions!
  - Course Information - CSE1500 Web and Database Technology (2025/26 Q2)

Literature DB

- **Books:**
  - **Fundamentals of Database Systems** Elmasri & Navathe
    - You can get quite far with just using the slide deck – the book is for deepening or more detailed or alternative explanations.
    - Most things we use can be found in the 6th edition, but there is a 7th edition which is a better investment
      - ISBN 1292097612 for 7th edition
      - Good book, very dry and theoretical though. Maybe buy it or find somebody who has it to share/ borrow. I guess you could survive without the book, though….
  - You do NOT need to buy the books below!
    - (We also do not really recommend you to buy them just for the sake of this course. We used these books for inspiration of this lecture, the important facts are in the slide set)
    - NoSQL and SQL Data Modeling: Bringing Together Data, Semantics, and Software
      - Ted Hills , 2016
      - ISBN 1634621093
    - SQL Antipatterns
      - Bill Karwin , 2010
      - ISBN 1934356557







Expectations

- This course is 5 ETCS
  - ~140 hours of work to pass
    - ~ 10 weeks, ~14 hours per week
  - During class weeks:
    - 4 hours of class and 4 hours of lab every week
      - 6 hours of self-study / homework?
        - » Or 10 if you do not go to the labs
    - No labs and class in exam prep weeks

What is the purpose of what?

- **Self-Study** (at home or in lab)
  - Study how things work in detail
    - e.g., how to normalize a schema step by step, how to build a B-tree step by step, etc.
    - Use all available material for that, e.g., lecture slides, transcripts, home study videos, books, internet resources & tutorials, YouTube videos, etc.
- **Classes**
  - **Editorialized view** on lecture content & core motivations and general approaches
    - e.g., why (or why not) would we want to normalize schemas?, why do we care about B-trees?, etc.
    - Note the program's main goal is socializing to develop an engineering mindset, not to focus on tiny technical details…
- **Labs**
  - Learn how to apply methodologies
  - Contains tutorials
  - "Do things on a computer"; "Experience the pain of fiddling with a DBMS"…
- **Practice Exams**
  - Danger of overfitting to previous exams….
    - Be careful with this!

Web Information Systems

Geert-Jan Houben
Full Professor

Alessandro Bozzon
(Affiliated) Full Professor

Marcus Specht
Full Professor @ CEL

Avishek Anand
Associate Professor

Christoph Lofi
Associate Professor

Claudia Hauff
Associate Professor

Maria Soledad Pera
Associate Professor

Asterios Katsifodimos
Assistant Professor

Fenia Aivaloglou
Assistant Professor @ CEL

Gosia Migut
Assistant Professor @ CEL

Jie Yang
Assistant Professor

Rihan Hai
Assistant Professor

Ujwal Gadiraju
Assistant Professor

Venktesh Viswanathan
Postdoctoral Researcher

Agathe Balayn
PhD Candidate

Ali Soleymani
PhD Candidate @ CEL

Alisa Rieger
PhD Candidate

Andra-Denis Ionescu
PhD Candidate

Anne, Arzberger
PhD Candidate

Arthur Barbosa Câmara
PhD Candidate

Christos Koutras
PhD Candidate

Danning Zhan
PhD Candidate

Esra de Groot
PhD Candidate

Gaole He
PhD Candidate

Garrett Allen
PhD Candidate

George (Georgios) Siachamis
PhD Candidate

Ioannis Petros Samiotis
PhD Candidate

Kyriakos Psarakis
PhD Candidate

Lijun Lyu
PhD Candidate

Lorenzo Corti
PhD Candidate

Manuel Valle Torre
PhD Candidate @ CEL

Nesse van der Meer
PhD Candidate @ CEL

Nirmal Roy
PhD Candidate

Peide Zhu
PhD Candidate

Philip Lippmann
PhD Candidate

Sara Salimzadeh
PhD Candidate

Shahin Sharifi
PhD Candidate

Shirley de Wit
PhD Candidate @ CEL

Shreyan Biswas
PhD Candidate

Tim Draws
PhD Candidate

Wenbo Sun
PhD Candidate

Xiaoling Zhang
PhD Candidate @ CEL

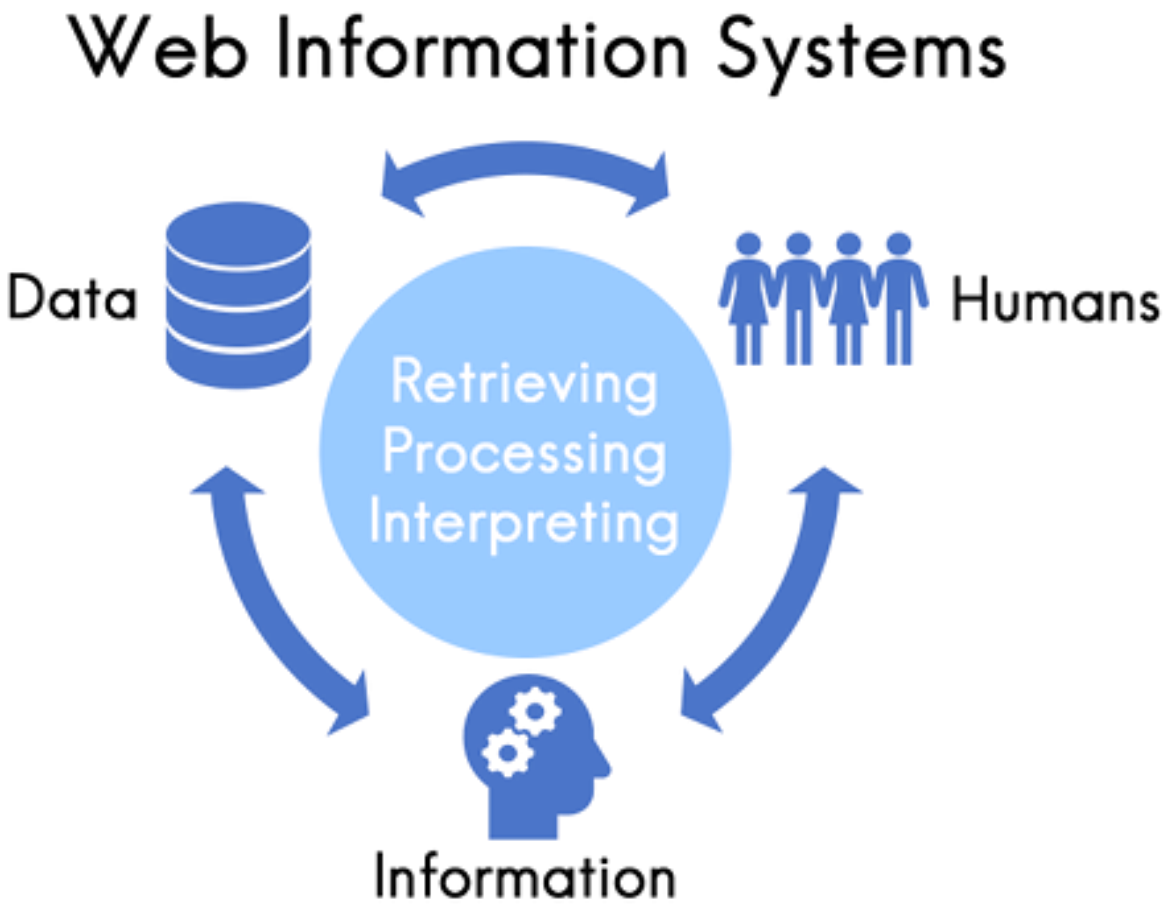
Yoon Lee
PhD Candidate @ CEL

Ziyu Li
PhD Candidate

Daphne
Secretary

Room - 820 West 4rd floor

Web Information Systems Research

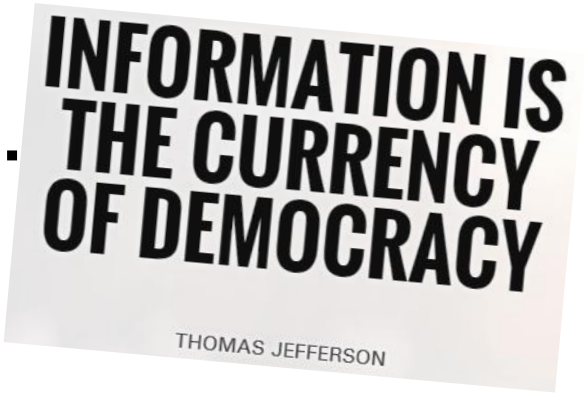


Information, Data and Knowledge?

“Isn’t that all the same…?”

Information and Data Modelling???

- Introduction to Information Theory…
  - … but what **is** information?

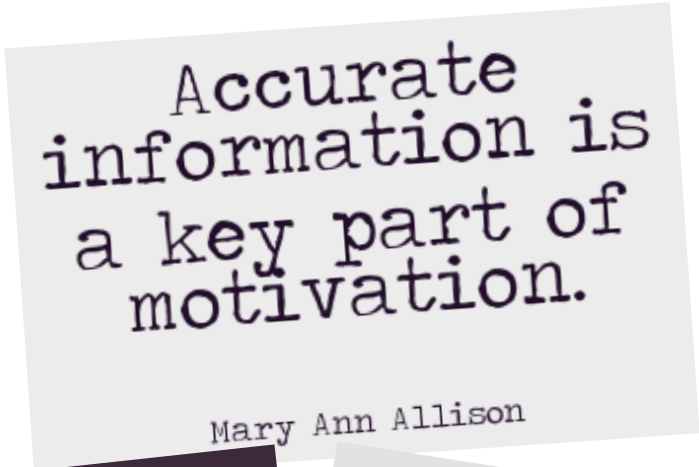


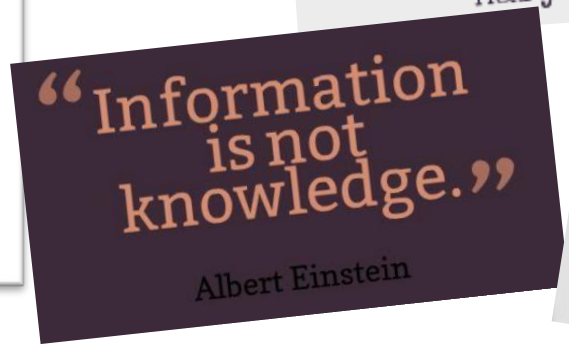
Information

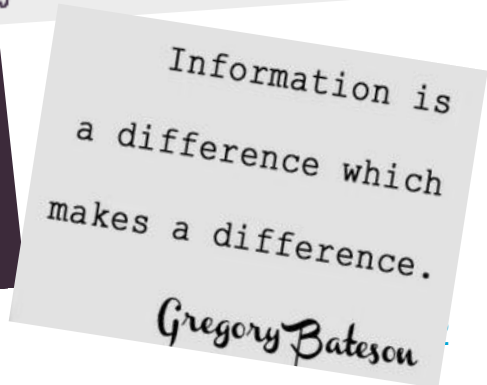
From Wikipedia, the free encyclopedia

For other uses, see Information (disambiguation).

Information (shortened as **info**) is that which informs. In other words, it is the answer to a question of some kind. It is also that from which data and knowledge can be derived, as data represents values attributed to parameters, and knowledge signifies understanding of real things or abstract concepts.[1] As it regards data, the information's existence is not necessarily coupled to an observer (it exists beyond an event horizon, for example), while in the case of knowledge, the information requires a cognitive observer.



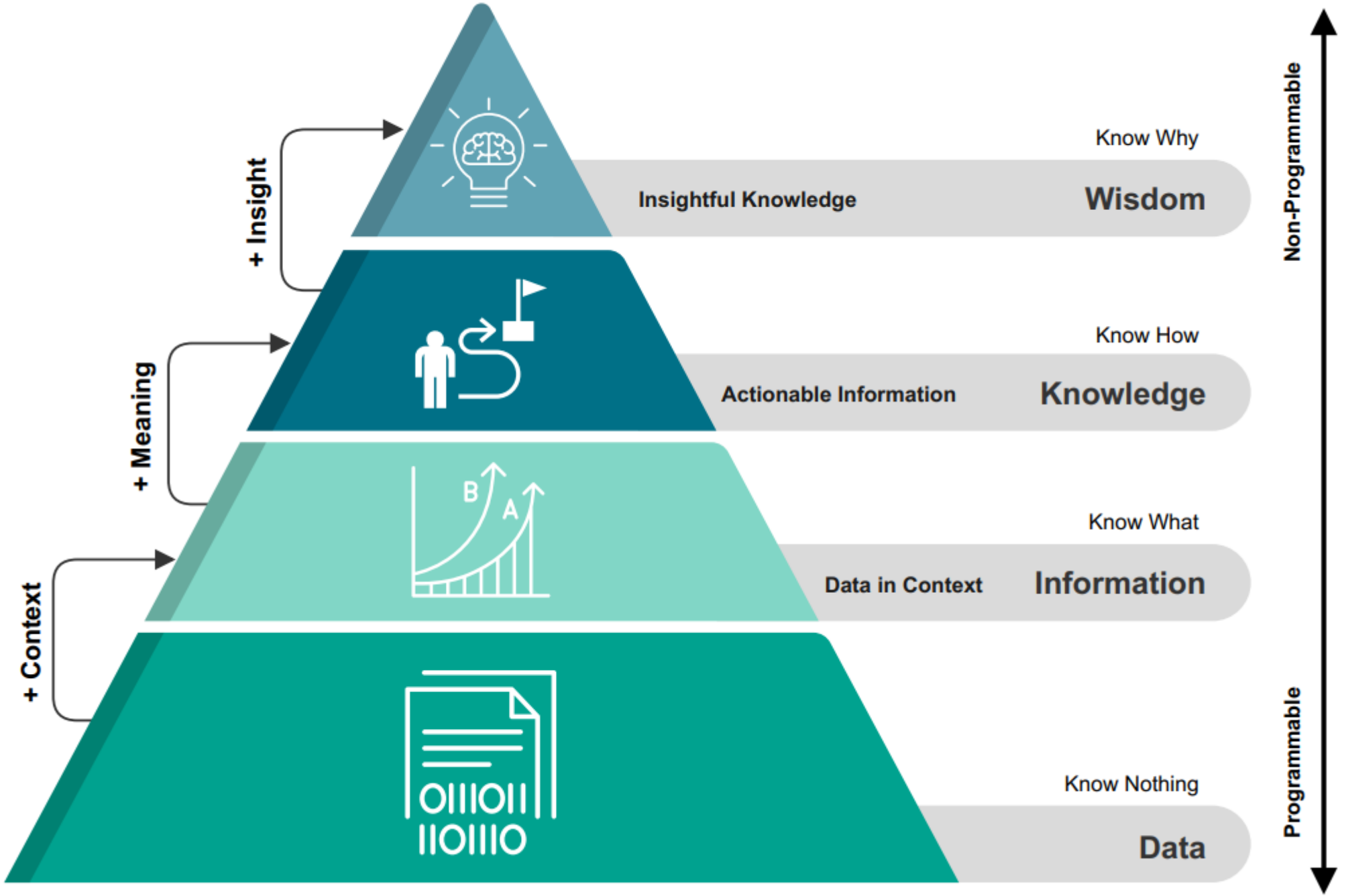




Information and Data Modelling???

- Another approach:
  - "Typically, **information** is defined in terms of data, **knowledge** in terms of **information**, and **wisdom** in terms of **knowledge**“
  - DIKW pyramid:
    - Not a formal model
    - Early versions liked to omit data, later versions like to omit wisdom

Rowley, Jennifer (2007). "The wisdom hierarchy: representations of the DIKW hierarchy".
Journal of Information and Communication Science. **33** (2): 163–180. doi:10.1177/0165551506070706.


+ Insight
+ Meaning
+ Context
Know Why
Insightful Knowledge
Wisdom
Know How
Actionable Information
Knowledge
Know What
Data in Context
Information
Know Nothing
Data
B
A
Non-Programmable
Programmable

The DIKW Pyramid

- Data as
  - … facts:
    - "as being discrete, objective facts or observations, which are unorganized and unprocessed and therefore have no meaning or value because of lack of context and interpretation.“
    - “raw facts”
  - … symbols:
    - "sets of signs that represent empirical stimuli or perceptions“
    - "recorded (captured or stored) symbols"
  - … signals
    - "signal readings“
    - "sensory stimuli, which we perceive through our senses"

- Zins, Chaim."Conceptual Approaches for Defining Data, Information, and Knowledge" (PDF). Journal of the American Society for Information Science and Technology. 58 (4): 479–493. 2007
- Rowley, Jennifer; Richard Hartley (2006). Organizing Knowledge: An Introduction to Managing Access to Information. Ashgate Publishing, Ltd. pp. 5–6. ISBN 978-0-7546-4431-6.
- Zins, Chaim (22 January 2007). "Conceptual Approaches for Defining Data, Information, and Knowledge". Journal of the American Society for Information Science and Technology. 58 (4): 479–493. doi:10.1002/asi.20508..

The DIKW Pyramid

- Aspects of Information
  - **Structural vs functional**
    - "organized or structured data, which has been processed in such a way that the information now has relevance for a specific purpose or context, and is therefore meaningful, valuable, useful and relevant."
  - **Symbolic vs subjective**
    - "the meaning of statements as they are intended by the speaker/writer”
    - "related to meaning or human intention“
  - **Information as Decrease Of Entropy**
    - Information Theory: Lower uncertainty on the state of the world

The DIKW Pyramid

- Aspects of Knowledge
  - “Knowledge is a fluid mix of **framed experience**, **values**, **contextual information**, **expert insight** and **grounded intuition** that provides an environment and framework for evaluating and incorporating new experiences and information.”
  - **Processed Knowledge**
    - "synthesis of multiple sources of information over time"
    - "a mix of contextual information, values, experience and rules"
  - **Procedural Knowledge**
    - "a mix of contextual information, expert opinion, skills and experience"
    - "information combined with understanding and capability"
  - **Propositional Knowledge**
    - “Knowledge is a thought in the individual's mind, which is characterized by the individual's justifiable belief that it is true. “

- Wallace, Danny P. (2007). Knowledge Management: Historical and Cross-Disciplinary Themes. Libraries Unlimited. pp. 1–14. ISBN 978-1-59158-502-2.

The DIKW Pyramid

- Wisdom….?
  - Very fuzzy…
    - “knowing the right things to do“
    - "integrated knowledge — information made super-useful“
    - "the ability to make sound judgments and decisions apparently without thought"

- Wallace, Danny P. (2007). Knowledge Management: Historical and Cross-Disciplinary Themes. Libraries Unlimited. pp. 1–14. ISBN 978-1-59158-502-2.
- Chisholm, James; Warman, Greg (2007). "Experiential Learning in Change Management". In Silberman, Melvin L. The Handbook of Experiential Learning. Jossey Bass. pp. 321–40. ISBN 978-0-7879-8258-4.

Example: Weather Forecast

- Data:
  - “Raw observations or facts without context.”
  - Temperature readings, humidity levels, wind speed, and barometric pressure at various locations.
- Information:
  - “Data organized and processed to provide context.”
  - A weather report that combines temperature, humidity, wind speed, and barometric pressure data to describe current weather conditions in a specific region.

Example: Weather Forecast

- Knowledge:
  - “Information that is interpreted and understood in a broader context.”
  - Meteorological analysis that explains the current weather patterns, predicts future weather conditions, and identifies the factors influencing the weather.
- Wisdom:
  - “Application of knowledge to make informed decisions.”
  - Using the meteorological knowledge to make decisions such as advising the public on potential weather hazards, helping farmers plan their planting seasons, or guiding businesses in making weather-sensitive decisions.

01 november • 15:49

Donderdag code geel in groot deel van het land wegens storm Ciarán

Morgen zal Nederland te maken krijgen met storm Ciarán. In de middag kunnen de kustgebieden windkracht 9 bereiken. Door de regen en sterke wind kunnen langere files dan normaal ontstaan. De ANWB adviseert...

The Modern World produces a humongous amount of Data

Quiz: Data Processing you Encountered Today?

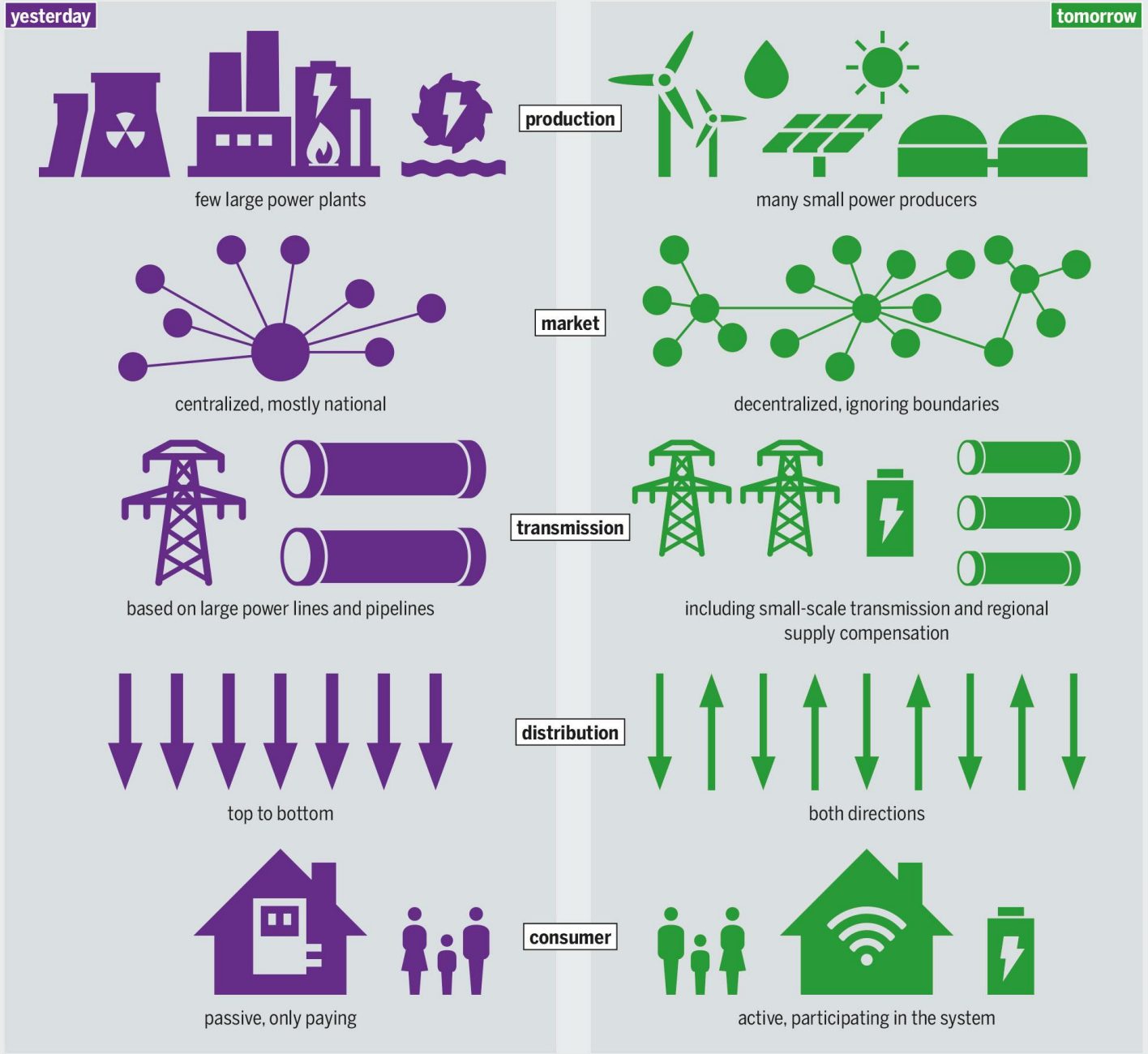
- Vevox Poll
  - https://tudelft.display.vevox.com#/present/819067/4MP39YJUS8O9JWA2V5XL
- https://tudelft.vevox.com/#/meeting/815129/polls

Amsterdam
Almelo
Roosendaal
Middelburg
Goes
Bergen
Helmond
A16
A58
57

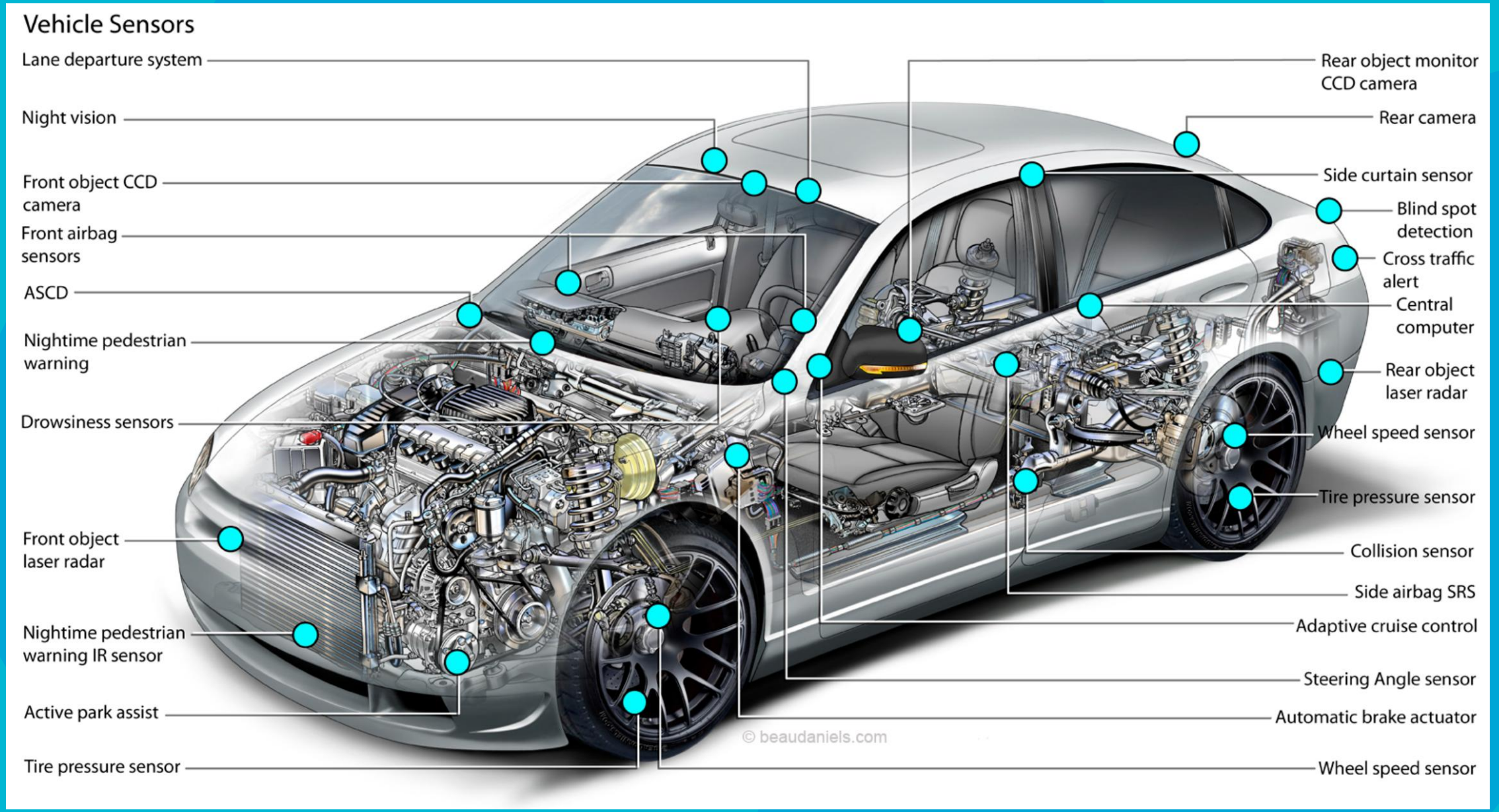
https://www.youtube.com/watch?v=qegdkl2GYrk

STAYING BIG OR GETTING SMALLER

Expected structural changes in the energy system made possible by the increased use of digital tools

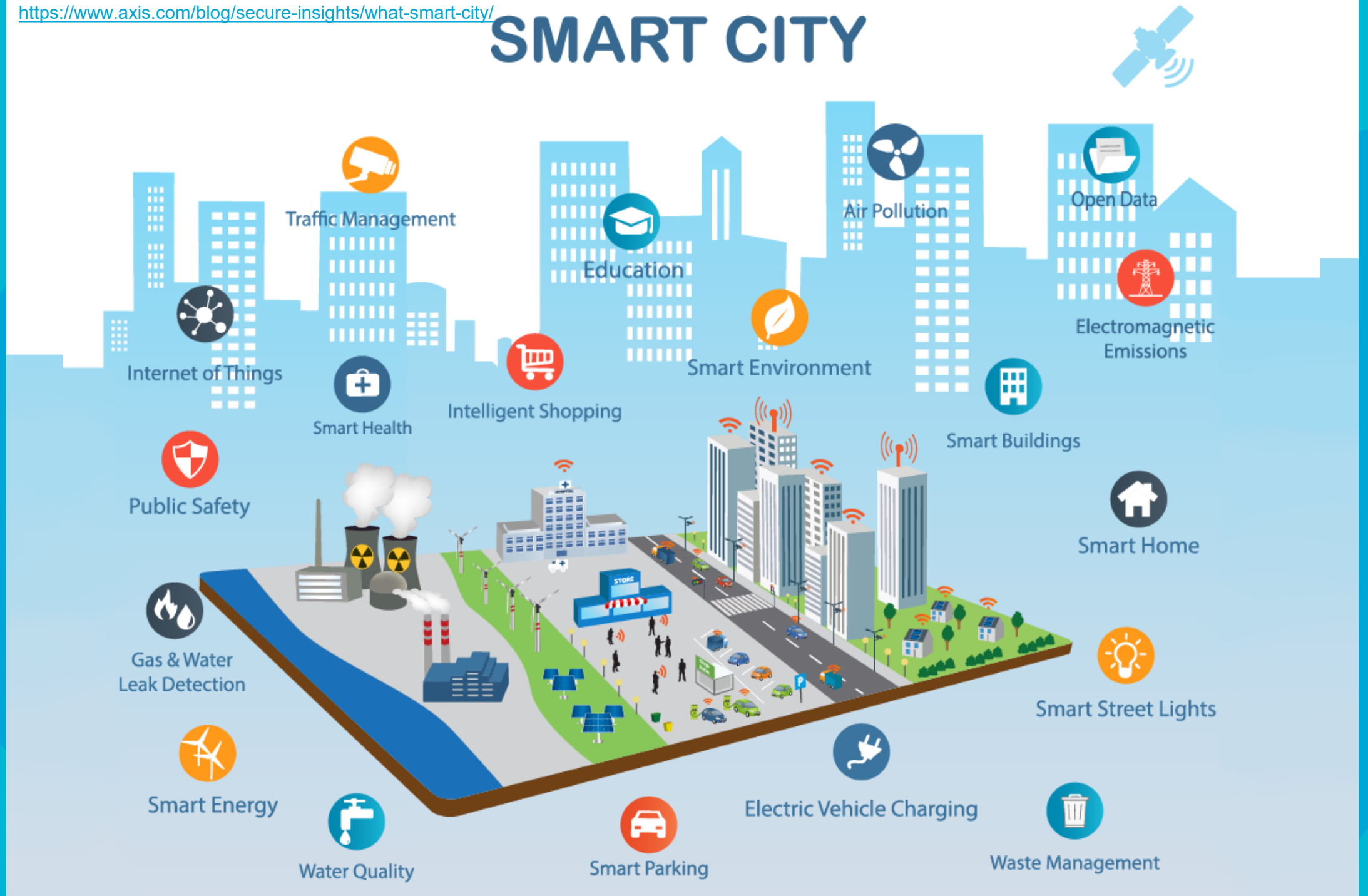


https://en.wikipedia.org/wiki/Smart_grid



https://beaudaniels.com/connected-car

https://www.axis.com/blog/secure-insights/what-smart-city/



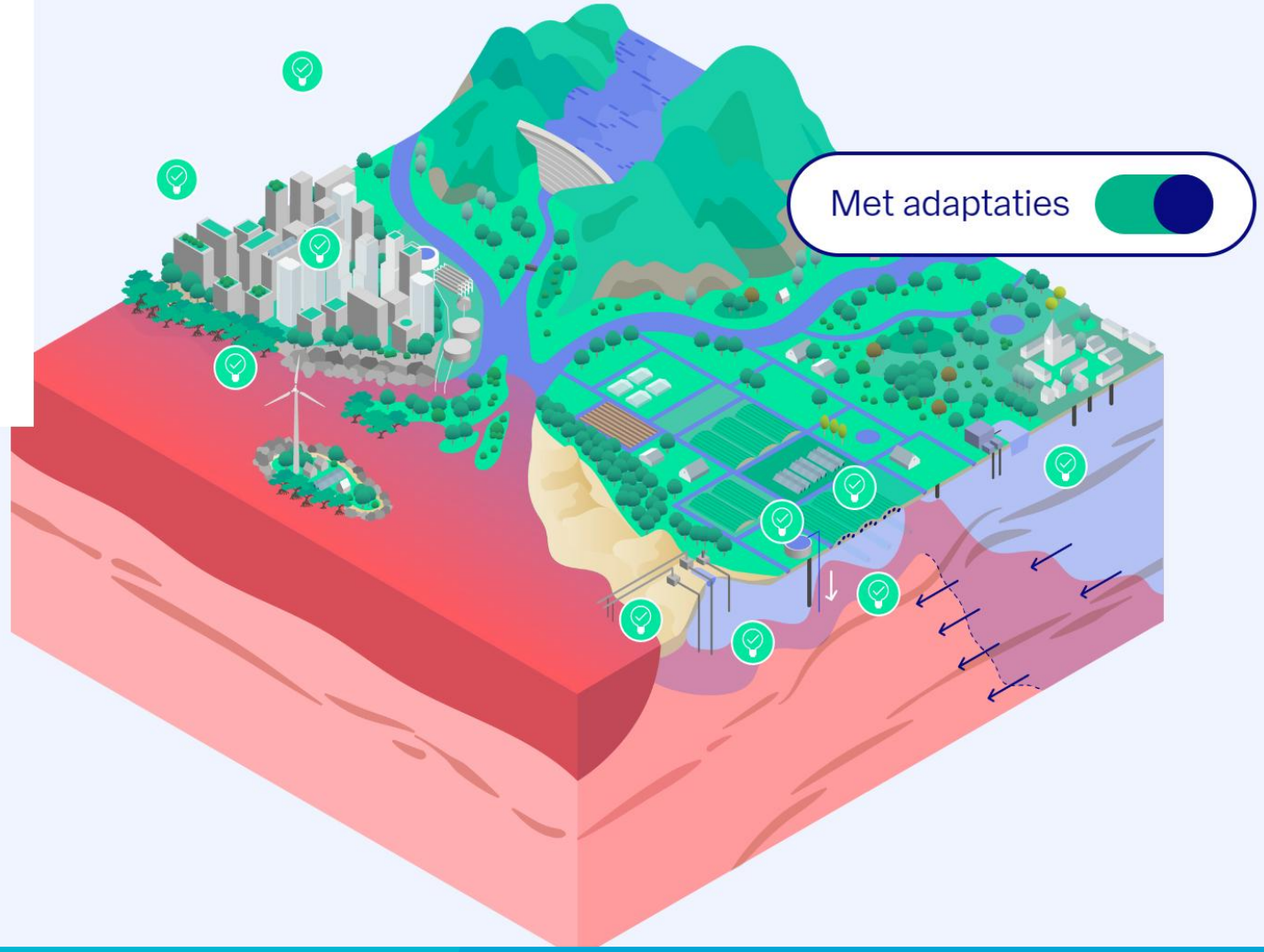
https://www.geovision.com.tw/

Om onze huidige levensstandaard en veiligheid te waarborgen moeten we ons aanpassen aan de nieuwe situatie.

Lukt het ons om de knop op tijd om te zetten?

Verzilting van het land in de tijd

De toekomst



https://deltalife.deltares.nl/september_2022/verzilting_de_trage_ramp

From Data to Information/Knowledge??
AI is here to fix it all!!!?

"Artificial intelligence (AI) will be the most powerful technology in generations for benefiting humanity."

"Artificial Intelligence (AI) has the potential to solve some of **society's most important challenges**, for example the mounting pressure on **healthcare**, while at the same time enriching and improving many other aspects of our society, such as **transport**. In short, the impact of AI is enormous, and it may even become the foundation of our future."

– from NWO LTP ROBUST

But…

- …how successful are we really with using AI?
  - …and especially with the ever-popular Deep Learning? (DL)
- DL is not magic - it is an incredibly powerful tool for extracting regularities from data according a given objective.
  - Claim #1: **A DL system will be just as smart as the data it gets.**
  - Claim #2: **A DL system will be just as smart as the objective it optimizes.**

Challenges

- Many open challenges…
  - **Where does data come from?**
  - **Data Quality**
    - The right training data?
    - Training Data Balanced?
  - **Model Evaluation**
  - **“Weird Features”**
  - **Common Sense Reasoning**
  - **Explanations**

From xkcd.com

"To err is human' but a human error is nothing to what a computer can do if it tries."

from Agathe Christie "Hallowe'en Party", 1969

"People worry that computers will get too smart and take over the world, but the real problem is that they're too stupid and they've already taken over the world."

from Pedro Domingos "The Master Algorithm", 2015

The “hidden depth” of ML

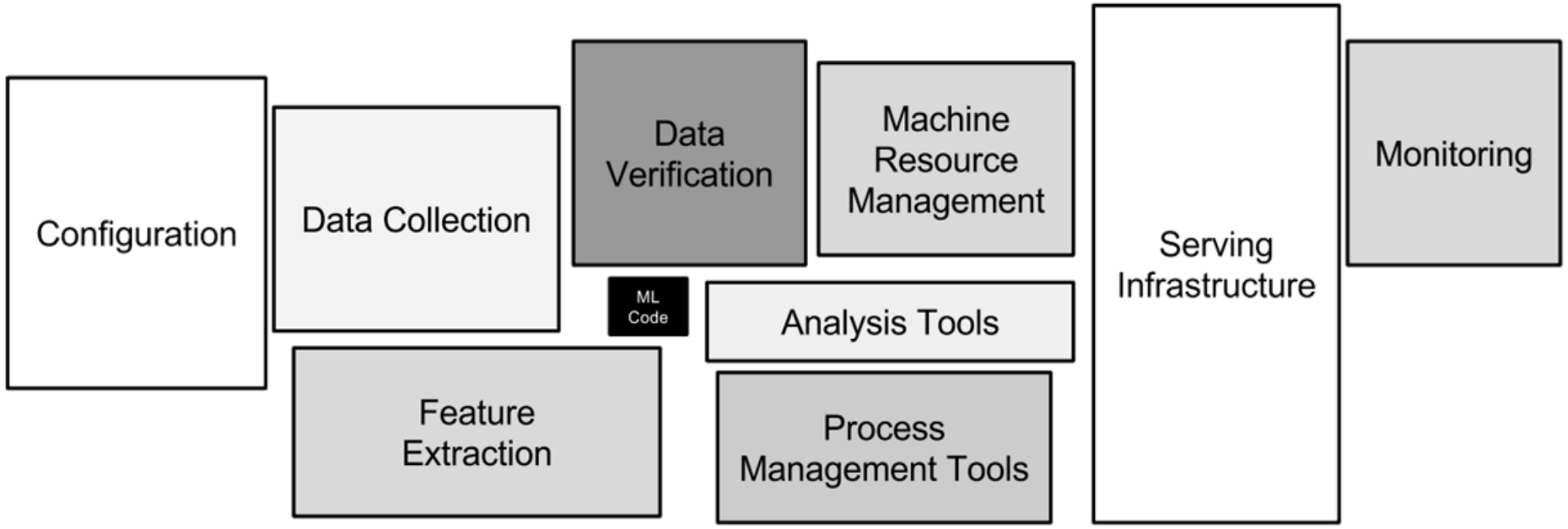
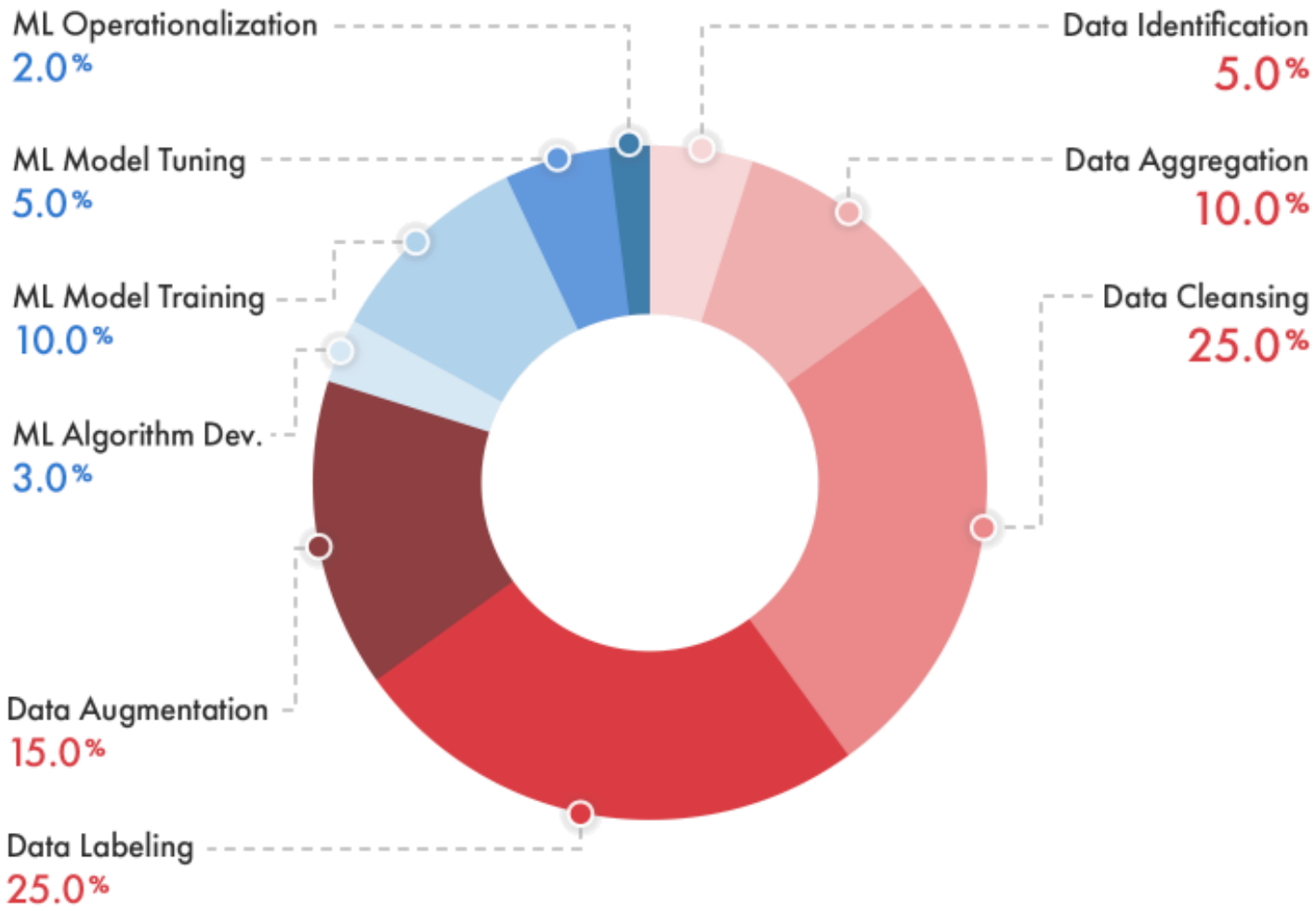


Figure 1: Only a small fraction of real-world ML systems is composed of the ML code, as shown by the small black box in the middle. The required surrounding infrastructure is vast and complex.

Hidden Technical Depth of Machine Learning, D. . Sculley, NIPS 2015

Percentage of Time Allocated to Machine Learning Project Tasks

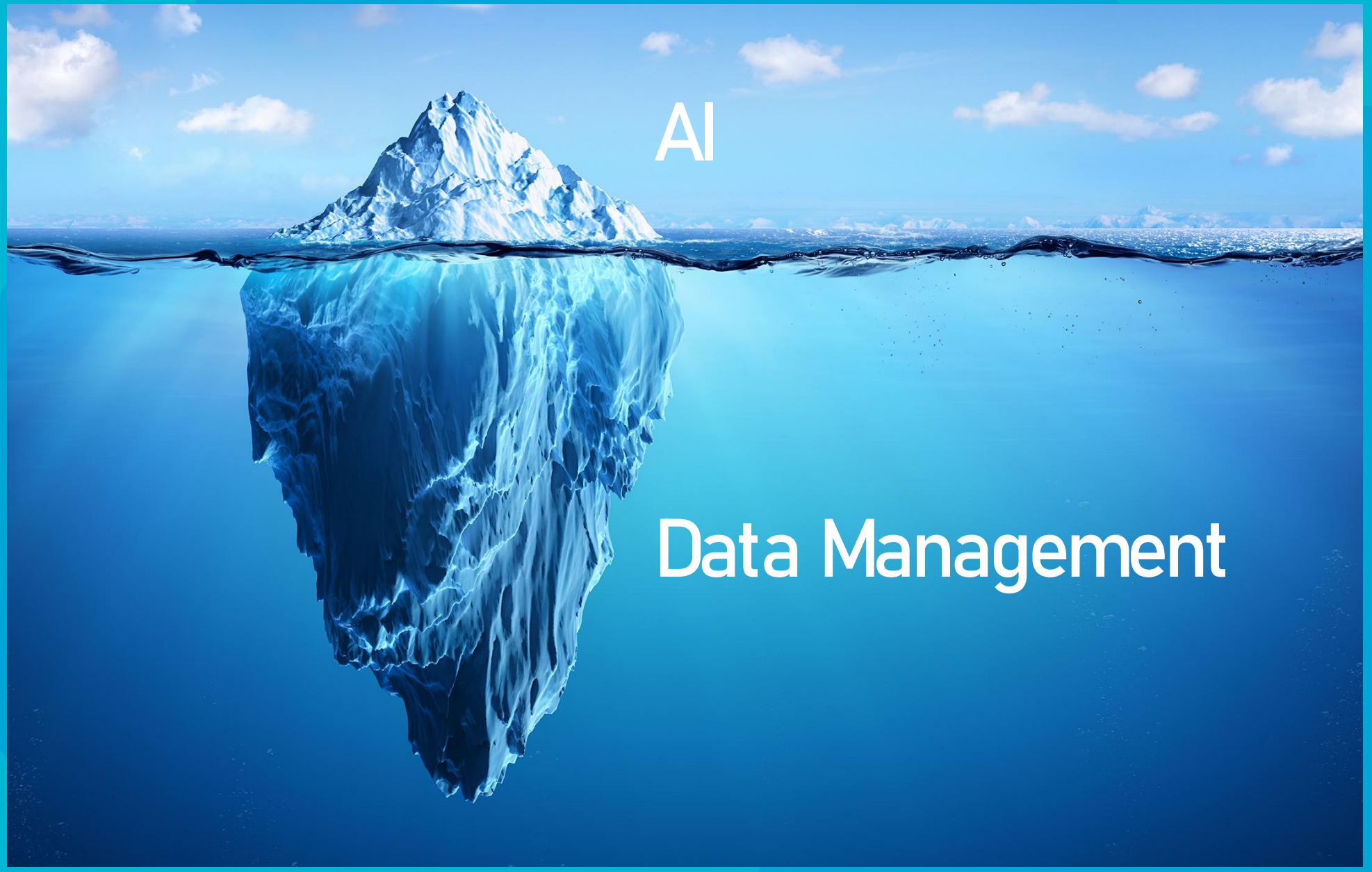


https://towardsdatascience.com/data-labeling-is-chinas-secret-weapon-in-the-connected-car-battle-e8e395965380

- https://tudelft.vevox.com/#/meeting/815129/polls

Into Databases!

TU Delft WIS Web Information Systems


AI
Data Management

TU Delft
WIS

Databases?

What Is a Database?

ORACLE®

Database defined

A database is an organized collection of structured information, or data, typically stored electronically in a computer system. A database is usually controlled by a database management system (DBMS). Together, the data and the DBMS, along with the applications that are associated with them, are referred to as a database system, often shortened to just database.

Data within the most common types of databases in operation today is typically modeled in rows and columns in a series of tables to make processing and data querying efficient. The data can then be easily accessed, managed, modified, updated, controlled, and organized. Most databases use structured query language (SQL) for writing and querying data.

https://www.oracle.com/database/what-is-database/#:~:text=Is%20a%20Database%3F-,Database%20defined,database%20management%20system%20(DBMS).

Databases?

Database

101 languages

Article Talk Read View source View history Tools

From Wikipedia, the free encyclopedia

This article is about the computing concept. For instances of the general concept, see Lists of databases.

In computing, a **database** is an organized collection of data or a type of data store based on the use of a **database management system** (**DBMS**), the software that interacts with end users, applications, and the database itself to capture and analyze the data. The DBMS additionally encompasses the core facilities provided to administer the database. The sum total of the database, the DBMS and the associated applications can be referred to as a **database system**. Often the term "database" is also used loosely to refer to any of the DBMS, the database system or an application associated with the database.

```
dvdrental=# select title, release_year, length, replacement_cost from film
dvdrental-#   where length > 120 and replacement_cost > 29.50
dvdrental-#   order by title desc;
          title          | release_year | length | replacement_cost
-------------------------+--------------+--------+------------------
 West Lion               |         2006 |    159 |            29.99
 Virgin Daisy            |         2006 |    179 |            29.99
 Uncut Suicides          |         2006 |    172 |            29.99
 Tracy Cider             |         2006 |    142 |            29.99
 Song Hedwig             |         2006 |    165 |            29.99
 Slacker Liaisons        |         2006 |    179 |            29.99
 Sassy Packer            |         2006 |    154 |            29.99
 River Outlaw            |         2006 |    149 |            29.99
 Right Cranes            |         2006 |    153 |            29.99
 Quest Mussolini         |         2006 |    177 |            29.99
 Poseidon Forever        |         2006 |    159 |            29.99
 Loathing Legally        |         2006 |    140 |            29.99
 Lawless Vision          |         2006 |    181 |            29.99
 Jingle Sagebrush        |         2006 |    124 |            29.99
 Jericho Mulan           |         2006 |    171 |            29.99
 Japanese Run            |         2006 |    135 |            29.99
 Gilmore Boiled          |         2006 |    163 |            29.99
 Floats Garden           |         2006 |    145 |            29.99
 Fantasia Park           |         2006 |    131 |            29.99
 Extraordinary Conquerer |         2006 |    122 |            29.99
 Everyone Craft          |         2006 |    163 |            29.99
 Dirty Ace               |         2006 |    147 |            29.99
 Clyde Theory            |         2006 |    139 |            29.99
 Clockwork Paradise      |         2006 |    143 |            29.99
 Ballroom Mockingbird    |         2006 |    173 |            29.99
(25 rows)
```

An SQL select statement and its result

Small databases can be stored on a file system, while large databases are hosted on computer clusters or cloud storage. The design of databases spans formal techniques and practical considerations, including data modeling, efficient data representation and storage, query languages, security and privacy of sensitive data, and distributed computing issues, including supporting concurrent access and fault tolerance.

Computer scientists may classify database management systems according to the database models that they support. Relational databases became dominant in the 1980s. These model data as rows and columns in a series of tables, and the vast majority use SQL for writing and querying data. In the 2000s, non-relational databases became popular, collectively referred to as NoSQL, because they use different query languages.

https://en.wikipedia.org/wiki/Database

Databases?

https://aws.amazon.com/what-is/database/

aws

What is a database?

A database is an electronically stored, systematic collection of data. It can contain any type of data, including words, numbers, images, videos, and files. You can use software called a database management system (DBMS) to store, retrieve, and edit data. In computer systems, the word *database* can also refer to any DBMS, to the database system, or to an application associated with the database.

Why is a database important?

A high-performing database is crucial to any organization. Databases support the internal operations of companies and store interactions with customers and suppliers. They also hold administrative information and more specialized data, such as engineering or economic models. Examples include digital library systems, travel reservation systems, and inventory systems. The following are some reasons why databases are essential.

Efficient scaling

Database applications can manage large amounts of data, scaling to millions, billions, and more. It's impossible to store this quantity of digital data without a database.

Data integrity

Databases often have built-in rules and conditions to maintain data consistency.

Data security

Databases support privacy and compliance requirements associated with any data. For example, to gain database access, users must log in. Different users might also have different levels of access, such as read-only.

Data analytics

Modern software systems use databases to analyze data. These systems can identify trends and patterns or make predictions. Data analytics help an organization make business decisions with confidence.

Databases?

https://www.techtarget.com/searchdatamanagement/definition/database

DEFINITION

database (DB)

By **Ben Lutkevich,** Technical Features Writer | **Adam Hughes**

What is a database?

A database is information that is set up for easy access, management and updating. Computer databases typically store aggregations of data records or files that contain information, such as sales transactions, customer data, financials and product information.

Databases are used for storing, maintaining and accessing any sort of data. They collect information on people, places or things. That information is gathered in one place so that it can be observed and analyzed. Databases can be thought of as an organized collection of information.

What are databases used for?

Businesses use data stored in databases to make informed business decisions. Some of the ways organizations use databases include the following:

- **Improve business processes.** Companies collect data about business processes, such sales, order processing and customer service. They analyze that data to improve these processes, expand their business and grow revenue.
- **Keep track of customers.** Databases often store information about people, such as

Database?

Let's define some terms!

- **Database**: "A collection of related data represented (using a data model and a defined data schema)"
- **Database Management System** (DBMS): "A software system managing and maintaining a database"

Let's define some terms!

- **Database**: represente… …ed data …l and a defined da…
- **Database**: “A softwar… …**n** (DBMS): …nd maintainin…

DB INTERNALS

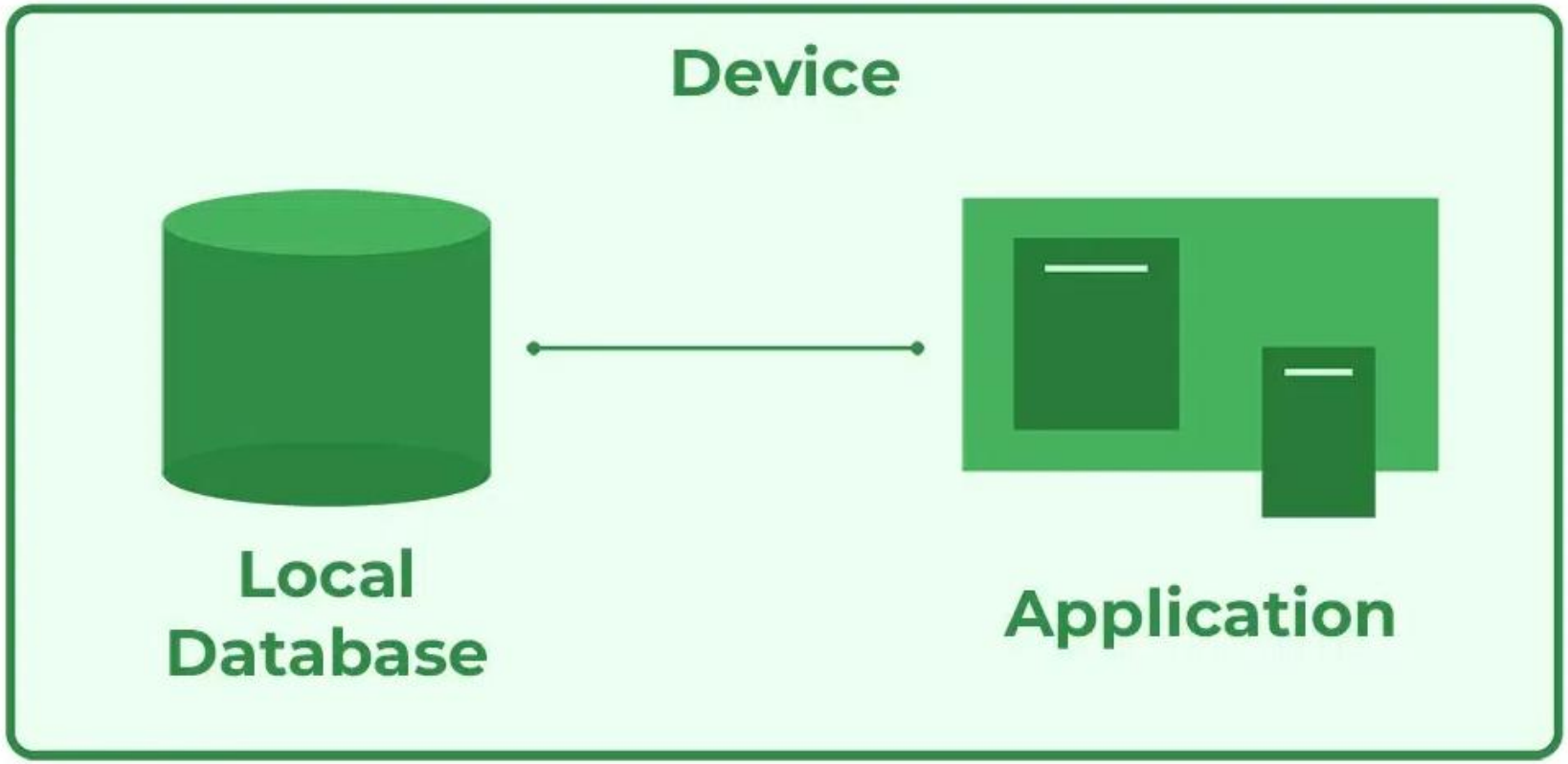
- ARCHITECTURE DIAGRAM

Why databases then (and not just files)?

- **1. Data Integrity & Consistency**
  - File systems lack built-in mechanisms for enforcing data rules.
  - Databases use constraints, transactions, and ACID properties to ensure reliable data
    - ACID Transactions: simplified “Complex operation will be executed reliably even with multiple uses or system failures”
- **2. Concurrency Control**
  - Multiple users accessing files can lead to conflicts or inconsistencies.
  - Databases handle concurrent access safely with locking and isolation levels.
- **3. Querying & Flexibility**
  - Databases offer powerful query languages (e.g., SQL) for fast, flexible data retrieval.
  - Languages are declarative: describe what you want, not how to get it!
- **4. Scalability & Performance**
  - File systems degrade with large data volumes and complex access patterns.
  - Databases are optimized for indexing, caching, and efficient storage.
- **5. Security & Access Control**
  - File systems offer basic permissions.
  - Databases provide fine-grained access control, authentication, and auditing.
- **6. Backup & Recovery**
  - Manual and error-prone in file systems.
  - Databases support automated backups, replication, and recovery tools.

Tier 1 Architecture

- Applications talk to local DB

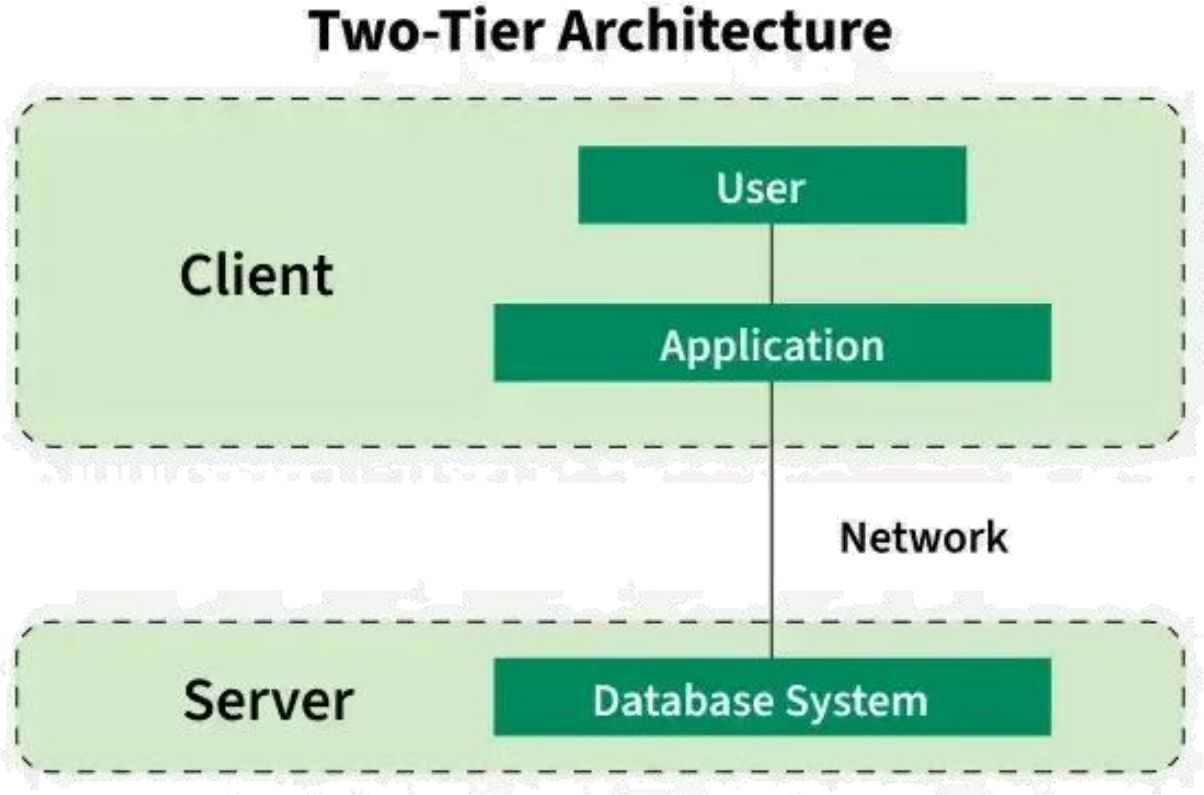


Tier 1: Architecture

- **Advantages of 1-Tier Architecture**
  - **Simple Architecture:** 1-Tier Architecture is the most simple architecture to set up, as only a single machine is required to maintain it.
  - **Cost-Effective:** No additional hardware is required for implementing 1-Tier Architecture, which makes it cost-effective.
  - **Easy to Implement:** 1-Tier Architecture can be easily deployed and hence it is mostly used in small projects.
- **Disadvantages of 1-Tier Architecture**
  - **Limited to Single User:** Only one person can use the application at a time. It's not designed for multiple users or teamwork.
  - **Poor Security:** Since everything is on the same machine, if someone gets access to the system, they can access both the data and the application easily.
  - **No Centralized Control:** Data is stored locally, so there's no central database. This makes it hard to manage or back up data across multiple devices.
  - **Hard to Share Data:** Sharing data between users is difficult because everything is stored on one computer.

Tier 2: Architecture

- Applications talk to shared DB on network

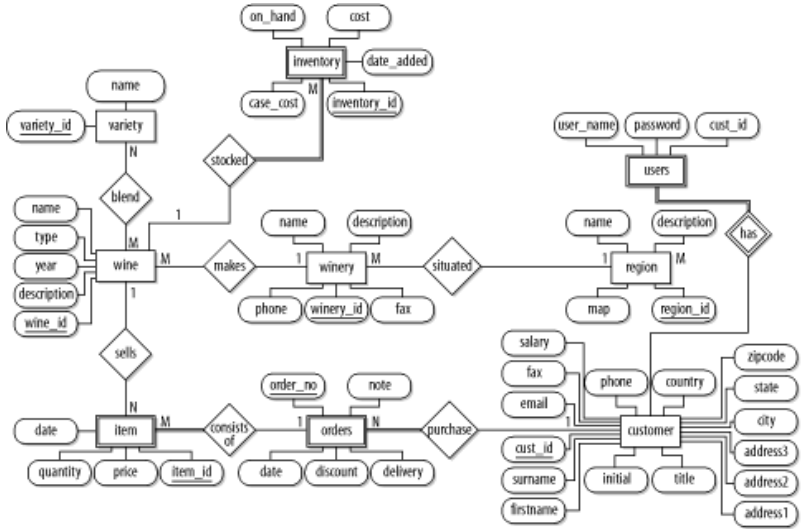


Tier 2 Architecture

- **Advantages of 2-Tier Architecture**
  - **Easy to Access:** 2-Tier Architecture makes easy access to the database, which makes fast retrieval.
  - **Scalable:** We can scale the database easily, by adding clients or upgrading hardware.
  - **Low Cost:** 2-Tier Architecture is cheaper than 3-Tier Architecture and
  - **Easy Deployment:** 2-Tier Architecture is easier to deploy than 3-Tier Architecture.
  - **Simple:** 2-Tier Architecture is easily understandable as well as simple because of only two components.
- **Disadvantages of 2-Tier Architecture**
  - **Limited Scalability:** As the number of users increases, the system performance can slow down because the server gets overloaded with too many requests.
  - **Security Issues:** Clients connect directly to the database, which can make the system more vulnerable to attacks or data leaks.
  - **Tight Coupling:** The client and the server are closely linked. If the database changes, the client application often needs to be updated too.
  - **Difficult Maintenance:** Managing updates, fixing bugs, or adding features becomes harder when the number of users or systems increases.

CSE1500 Overview

- **Week 1**
  - Lecture: **Introduction to Database Systems**
  - Lecture: **Database Modelling**
  - Handout Assignment: DB01-Schemas



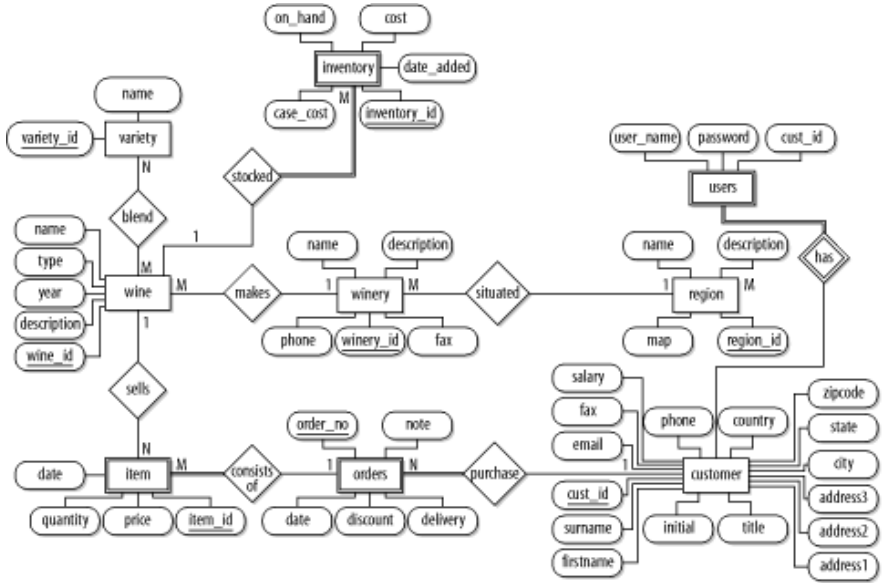
PostgreSQL

IBM DB2

ORACLE

CSE1500 Overview

- **Week 2**
  - Lecture: **Jumpstart SQL Query Language**
  - Lecture: **Database Modelling (part 2)**
  - Handout Assignment: DB02-SQL
  - Deadline Assignment: DB01-Schemas



CSE1500 Overview

- **Week 3**
  - Lecture: **Intermediate SQL Query Language**
  - Lecture: **Data Models & The Relational Model**
  - Handout Assignment: DB03-Relational Model
  - Deadline Assignment: DB02-SQL

CSE1500 Overview

- **Week 4**
  - Lecture: **The Relational Model, Functional Dependencies & Normalizations**
  - Lecture: Roundup & Q/A
  - Deadline Assignment: DB02-SQL
- **Week 5**
  - **Midterm Exam: Database Technology**

CSE1505 Overview

- CSE1505
  - **Internals of Relational Databases**
    - **Traditional Implementation**
    - Physical Storage & Physical Models
    - Query Processing & Query Optimization
    - Transaction Processing
  - **NoSQL Systems**
    - Conceptual Difference
    - Usage Scenarios

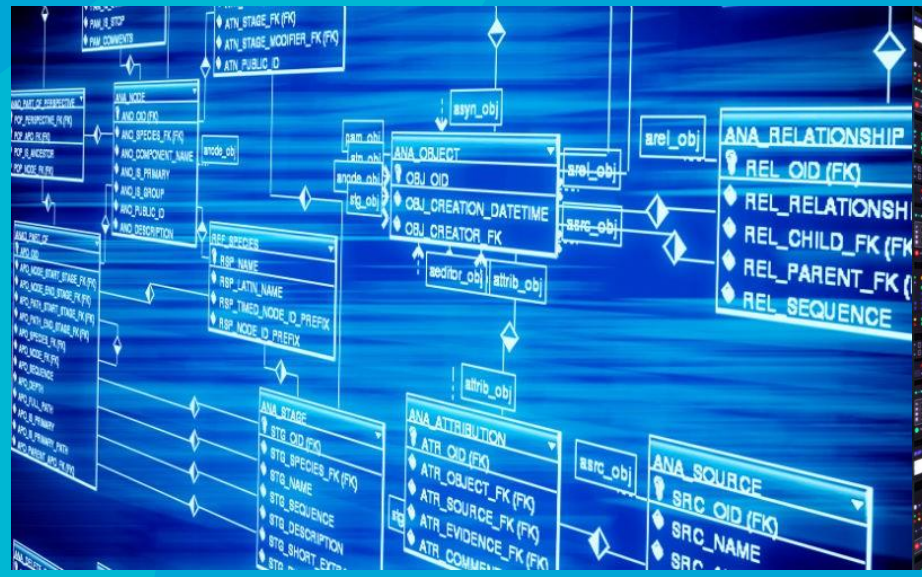
Goal of this course cycle…

- **BSc: Web and Database Technologies**
  - Focus on how to USE databases
    - SQL
    - Play around with different products
  - Learn how to make (simple!) Web applications
- **BSc: Information and Data Management**
  - Focus on how (relational) databases work
  - Theory, concepts, and foundations
- **BSc: Machine Learning**
- **BSc Electives: Big Data Processing, Data Mining, Computational Intelligence**
- **MSc Lectures**
  - Specialized topics like scalability, streams, Web, crowd, Data & AI, etc.

Summary

- In CSE1500/CSE1505, we care about
  - Well-structured schemas, conceptional and logical
    - Incl controlling data redundancy
  - “Independence” of data and application
  - Powerful Declarative Query Languages like SQL
  - Query Optimization and Access Path Optimization
    - e.g., incl Indexing
  - Transaction Management and Recovery

Towards Modelling!



TU Delft WIS Web Information Systems

Let's define some terms!

- **Database**: "A collection of related data represented (using a data model and a defined data schema)"
- **Database Management System** (DBMS): "A software system managing and maintaining a database"
- **Data Model**: "A formal definition on how to represent data (in general) and the available data operations"
- **Data Schema**: "A definition of the structure of a specific database"

Example: Database

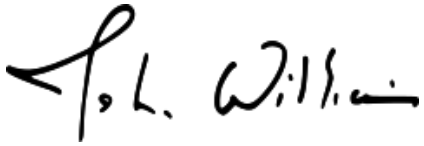
- Make a music encyclopedia **database**!
  - "A collection of related data represented (using a data model and a defined data schema)"
- Like musicbrainz.org
- Capture data on artists, their recordings, concerts, albums, etc.

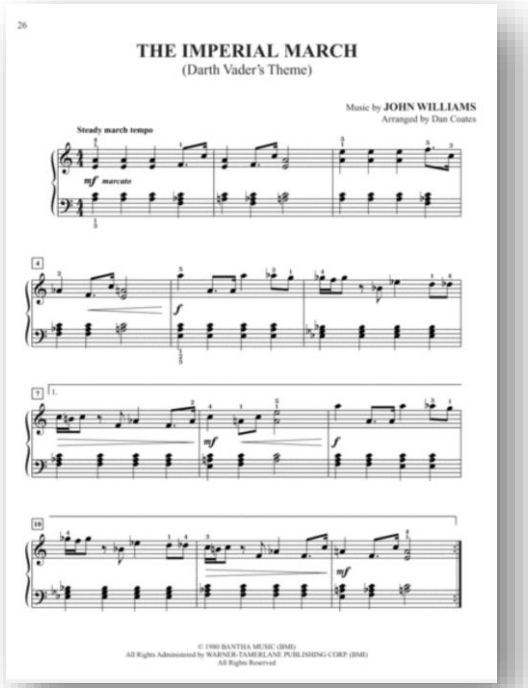
Image from: https://www.publicdomainpictures.net/en/view-image.php?image=194450&picture=background

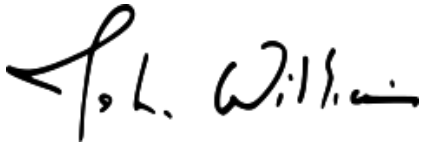
Example: Conceptual Schemas

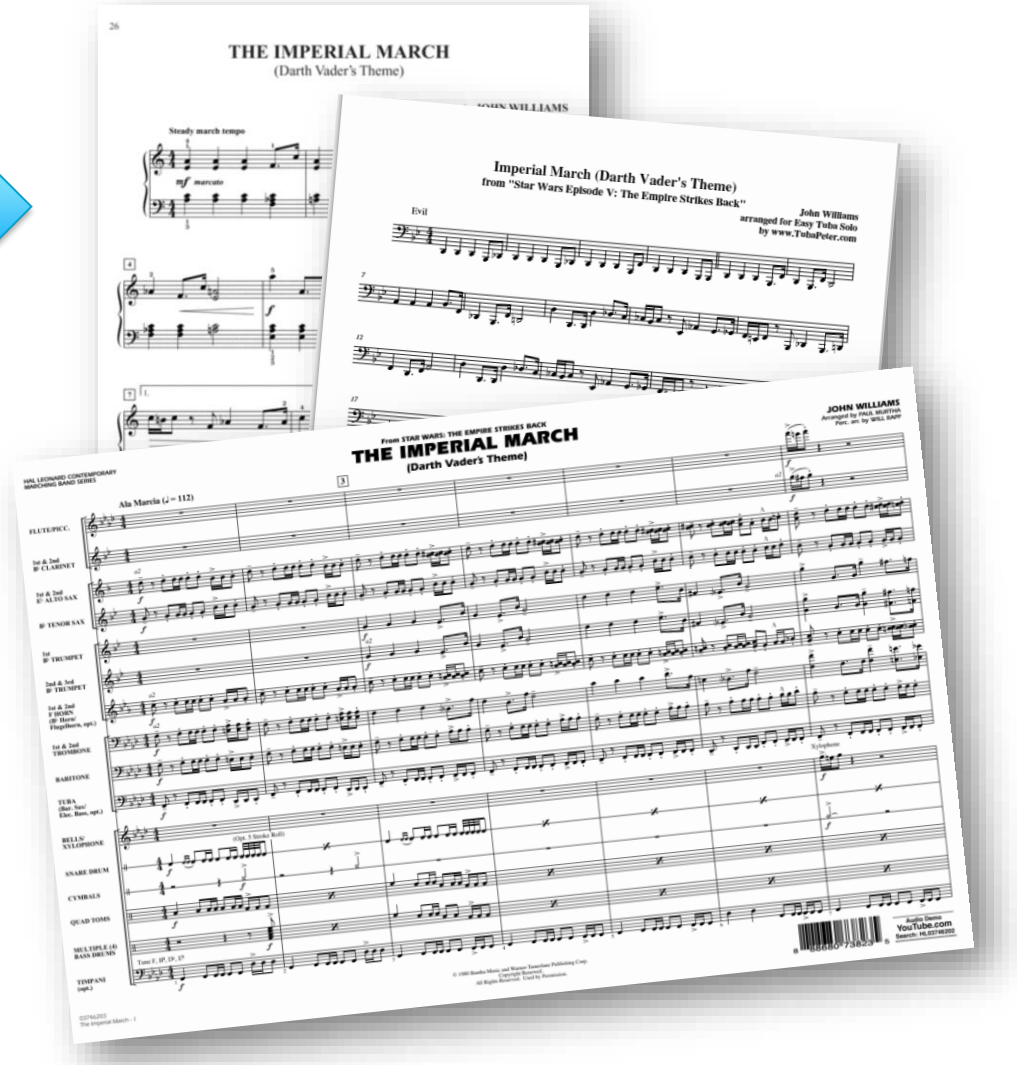
- We care about two types of **schemas**: conceptual and logical
  - **Schema:** "A definition of the structure of a specific database"

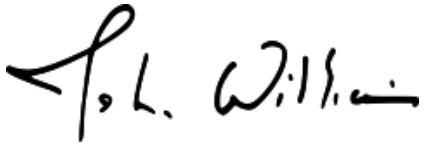
- **Conceptual Schemas:** Describes all entity types and relationships types to be stored in a database
  - "What do I want to store? (and what not?)
  - **Entity Type**: A type of "thing" which exists in the real world
    - Like "artist", "recording", "album", etc.
  - **Relationships**: "Connections" between entities
    - Like "an artist *records* a recording", "a recording is *contained in* an album", etc.

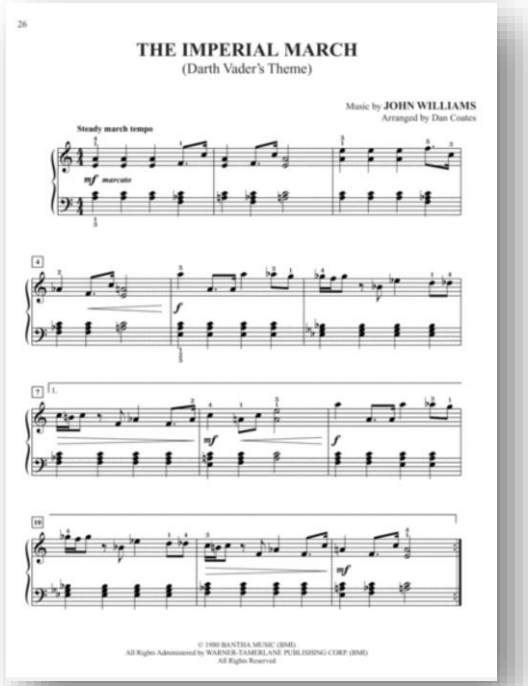



26
THE IMPERIAL MARCH
(Darth Vader's Theme)
Music by JOHN WILLIAMS
Arranged by Dan Coates
Steady march tempo
mf marcato
f
mf
f
mf
1.
© 1980 BANTHA MUSIC (BMI)
All Rights Administered by WARNER-TAMERLANE PUBLISHING CORP. (BMI)
All Rights Reserved



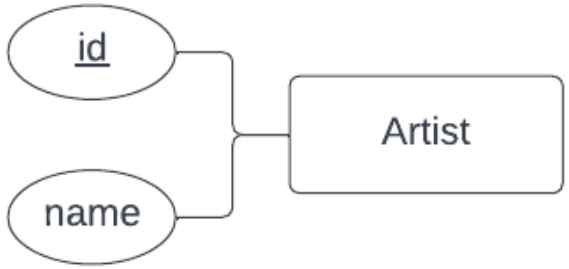

THE IMPERIAL MARCH
(Darth Vader's Theme)
Steady march tempo
Imperial March (Darth Vader's Theme)
from "Star Wars Episode V: The Empire Strikes Back"
John Williams
arranged for Easy Tuba Solo
by www.TubaPeter.com
Evil
From STAR WARS: THE EMPIRE STRIKES BACK
THE IMPERIAL MARCH
(Darth Vader's Theme)
JOHN WILLIAMS
Arranged by PAUL MURTHA
Perc. arr. by BILL BARD
HAL LEONARD CONTEMPORARY MARCHING BAND SERIES
Alla Marcia (♩= 112)
FLUTE/PICC.
1st & 2nd B♭ CLARINET
1st & 2nd E♭ ALTO SAX
B♭ TENOR SAX
1st B♭ TRUMPET
2nd & 3rd B♭ TRUMPET
1st & 2nd F HORN (B♭ Horn*, Flugelhorn, opt.)
1st & 2nd TROMBONE
BARITONE
TUBA (Bar. Sax/ Elec. Bass, opt.)
BELLS/ XYLOPHONE
SNARE DRUM
CYMBALS
QUAD TOMS
MULTIPLE (4) BASS DRUMS
TIMPANI (opt.)
Xylophone
Audio Demo
YouTube.com

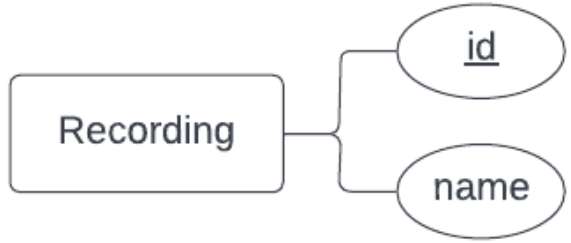


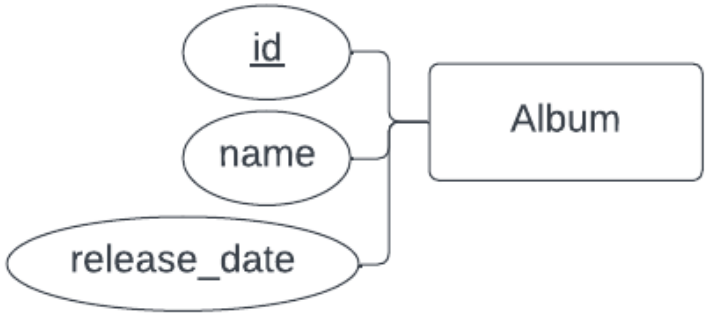

THE IMPERIAL MARCH
(Darth Vader's Theme)
Music by JOHN WILLIAMS
Arranged by Dan Coates
Steady march tempo


Star Wars: The Empire Strikes Back (Original Motion Picture Soundtrack)
THE ORIGINAL MOTION PICTURE SOUNDTRACK
STAR WARS
A NEW HOPE
JOHN WILLIAMS
THE STAR WARS TRILOGY

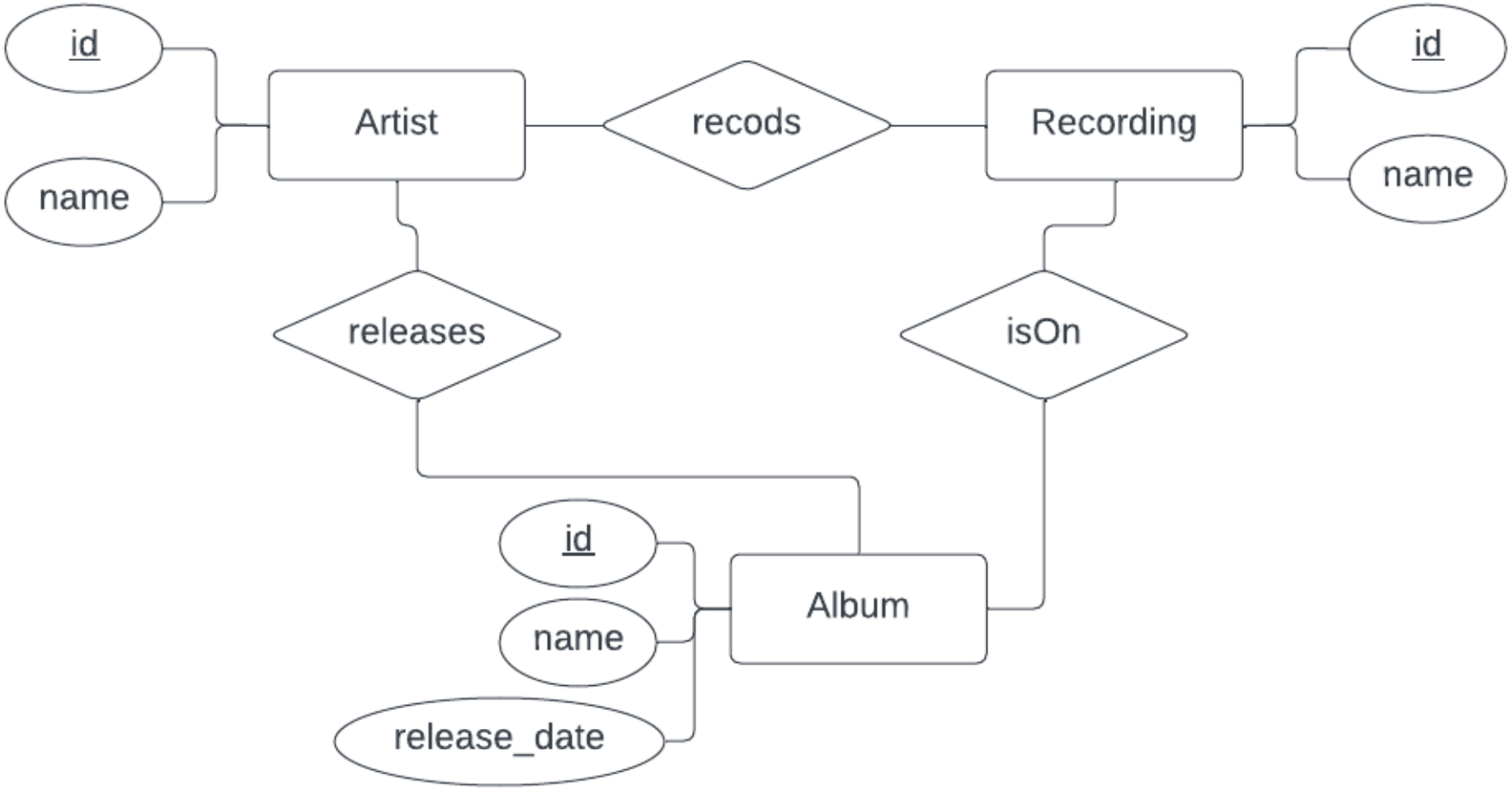
Example Conceptual Schemas:
What Entity Types do we care about?





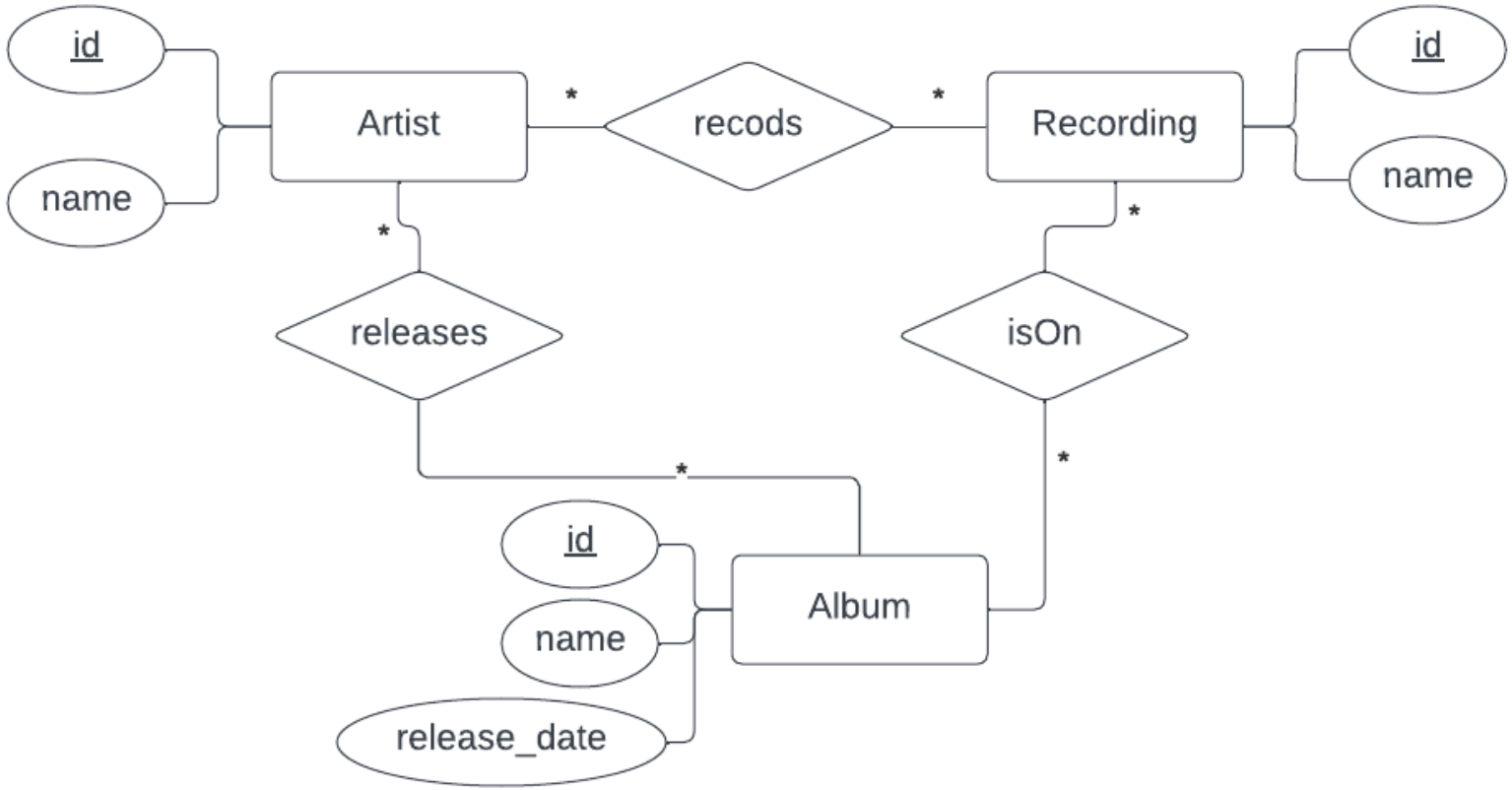


Example Conceptual Schemas: What Relationships do we care about?



This is an **ER Diagram** (Entity – Relationship). It shows Entities (Types) and their Relationships (Types).

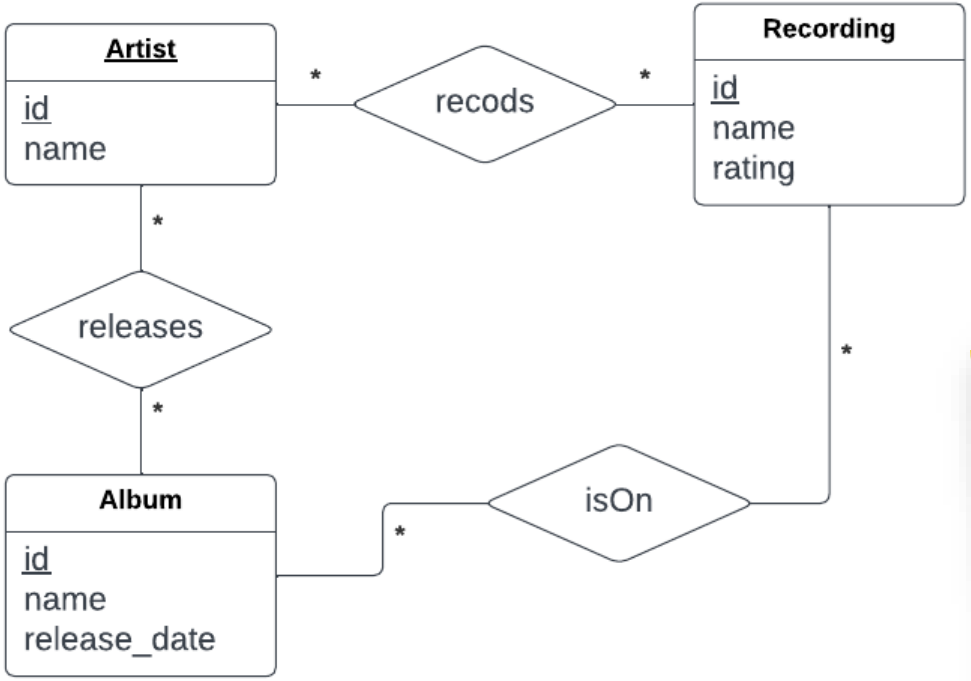
Example Conceptual Schemas: What Cardinalities do the relationships have?

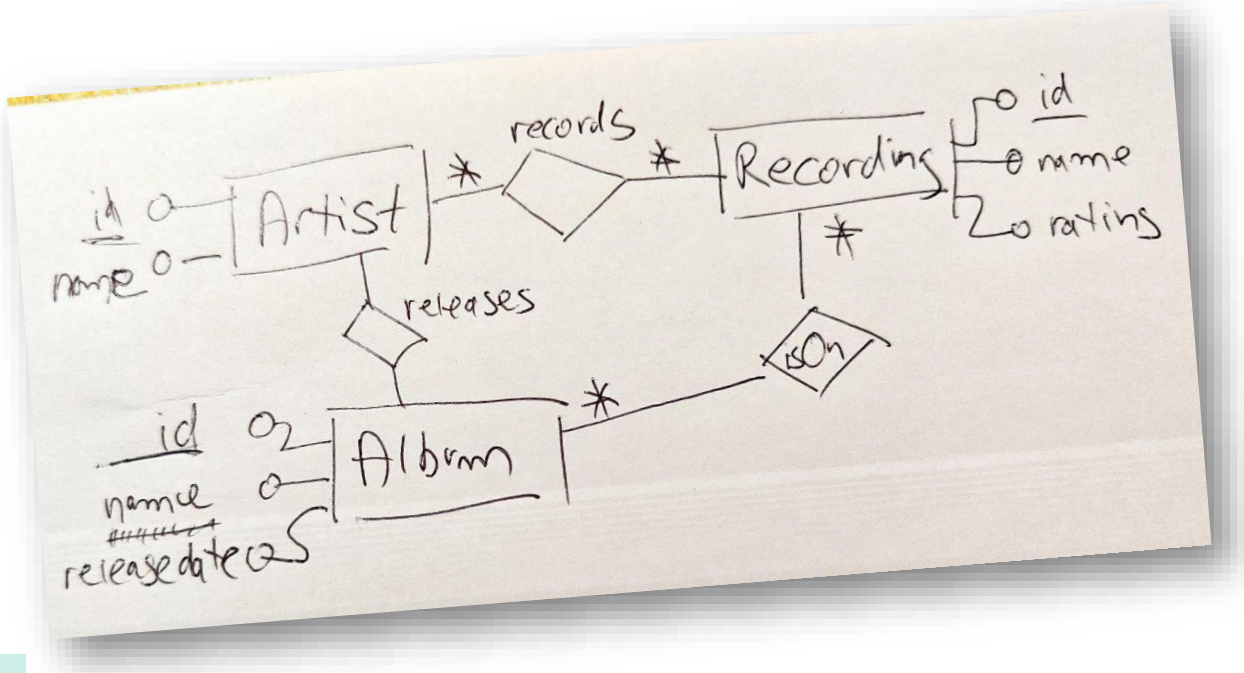


The “n-m” here just means just stands for “many-to-many”:
Recording can belong to many albums; an album can have many recordings.
Often, people also use “*” instead of “n” or “m”.

This diagram style (with circles for attributes, boxes for entities, diamonds for Relationships” is called “Chen-style ER Entity-Relationship Diagrams”. This is the traditional style for representing ER diagrams.

Example Conceptual Schemas: Alternative Diagram Styles





UML-Style Entity Types. Works well with software, annoying when drawing by hand…
I guess this is what most people would want to use when using schema diagramming software.

Sloppy handwritten Chen ER-inspired diagram with tiny circles for attributes; and incomplete cardinalities (minus points in exam!)
A pretty version of this is used later in this slide deck.

Data Models / Theory

- **Data Model**: “A formal definition on how to represent data (in general) and the available data operations.”

- A **data model** consists of three parts
  - Structure
    - **Data structures** are used to create databases representing the modeled objects
  - Integrity
    - Rules expressing the **constraints** placed on these data structures to ensure structural integrity
  - Manipulation
    - Operators that can be applied to the data structures, to **update** and **query** the data contained in the database

Example Data Model: Relational Model

- **Data Model**: “A formal definition on how to represent data (in general) and the available data operations.”

- The **Relational (Data) Model**
  - Represents multiple data **tables** which are linked to each other (via ids and foreign keys)
  - Typically: Each entity type and relationship type has its own table
    - A Database Management System (DBMS) manages those tables and their schemas (like PostgreSQL, Oracle, MySQL, etc.)
  - Powerful SQL Queries can be used to construct any result we need
    - The DBMS makes sure to execute such queries efficiently
  - Most large web platforms/enterprise backends use relational DBs

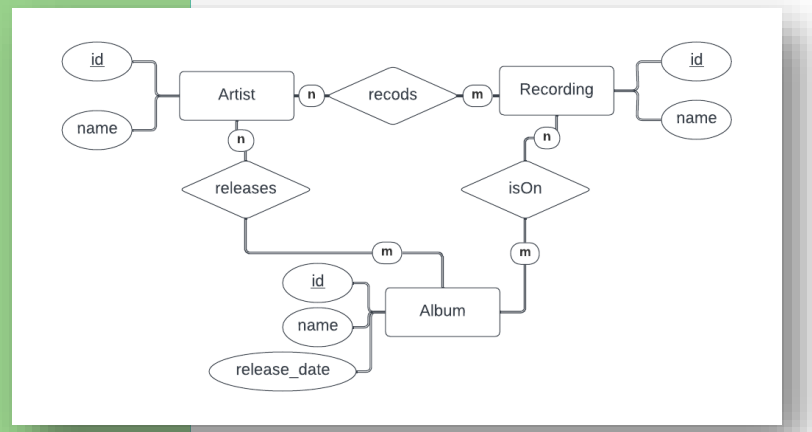
Example: Logical Schemas

- We care about two types of **schemas**: conceptual and logical
  - **Schema:** "A definition of the structure of a specific database"

- **Logical Schemas:** Represent all entity types and relationships from a conceptual schema in the chosen data model
  - Conceptual schema was independent of data model
    - For the logical schema, we adapt the conceptual schema with the capabilities and restrictions of the data model

Example: A logical schema of our music database using the relational model

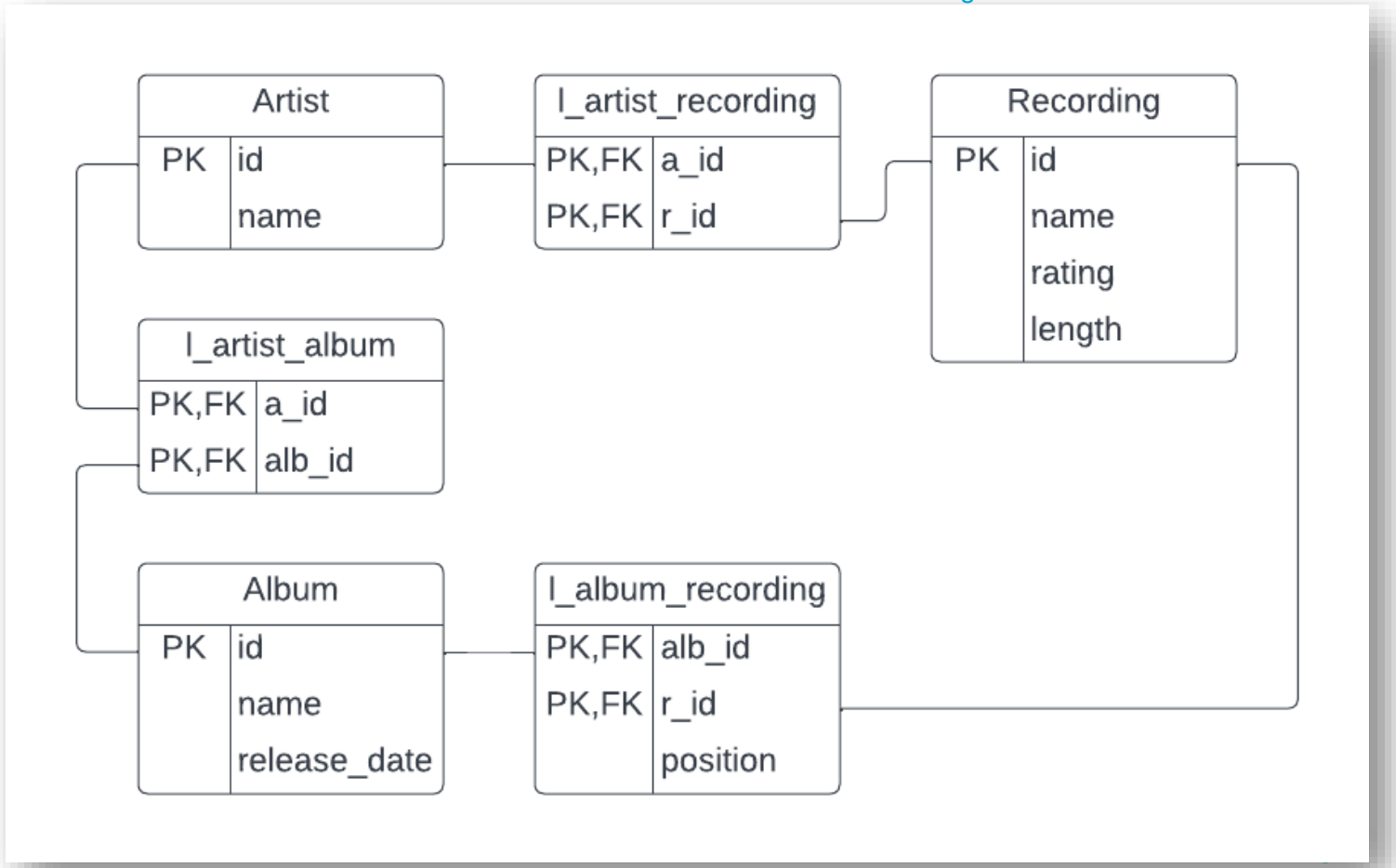
Conceptual Schema



Focus on Entity Types and their relationships

Logical (Relational) Schema

Focus the resulting relations / tables and constraints



Artist

id	name
85	Madness
86	Qaballah Steppers
87	Steve Rossiter
88	Enya
89	Madonna
91	Bill Cosby
92	Yanni
93	Metallica
94	John Williams
96	The Bangles
98	Burl Ives
99	Bing Crosby
101	Frank Sinatra
104	[Disney]
110	Harry Connick, Jr.
111	Mariah Carey
116	Enigma
119	George Winston
121	The Chieftains
122	New Edition
123	Cliff Richard
125	Dan Fogelberg
127	James Taylor
129	Simon & Garfunkel

l_artist_recording

a.id	r.id
94	257208
94	334474
94	447670
94	453112
94	453118
94	453119
94	464602
94	464603
94	464604
94	496934
94	515742
94	515746
94	515748
94	532750
94	532751
94	532752
94	532754
94	532755
94	532757

Recording

record_id	record_name	rating...	length
257208	Superman: Main Theme	4	04:26
334474	Prologue	6	02:12
447670	Duel of the Fates	7	04:14
453112	Opening Titles	8	00:33
453118	Welcome to Jurassic Park	8	07:54
453119	My Friend, the Brachiosaurus	6	04:16
464602	Ah, Rats!!!	8	03:41
464603	Escape From Venice	6	04:24
464604	No Ticket	6	02:46
496934	E.T.'s New Home	4	01:38
515742	Yoda's Theme	6	03:29
515746	Yoda and the Force	6	04:02
515748	Lando's Palace	6	03:53
532750	Star Wars Main Title / Ambush on Corus…	8	03:46
532751	Across the Stars (Love Theme From Atta…	8	05:33
532752	Zam the Assassin / The Chase Through C…	10	11:07
532754	Departing Coruscant	6	01:44
532755	Anakin and Padmé	6	03:56
532757	The Meadow Picnic	8	04:14
532758	Bounty Hunter's Pursuit	8	03:23
532759	Return to Tatooine	8	06:56
532760	The Tusken Camp / The Homestead	6	05:54
532761	Love Pledge / The Arena	8	08:29
532762	Confrontation With Count Dooku / Finale	8	10:45

Artist

id	name
85	Madness
86	Qaballah Steppers
87	Steve Rossiter
88	Enya
89	Madonna
91	Bill Cosby
92	Yanni
93	Metallica
94	John Williams
96	The Bangles
98	Burl Ives
99	Bing Crosby
101	Frank Sinatra
104	[Disney]
110	Harry Connick, Jr.
111	Mariah Carey
116	Enigma
119	George Winston
121	The Chieftains
122	New Edition
123	Cliff Richard
125	Dan Fogelberg
127	James Taylor
129	Simon & Garfunkel

l_artist_recording

a.id	r.id
94	257208
94	334474
94	447670
94	453112
94	453118
94	453119
94	464602
94	464603
94	464604
94	496934
94	515742
94	515746
94	515748
94	532750
94	532751
94	532752
94	532754
94	532755
94	532757

Recording

record_id	record_name	rating...	length
257208	Superman: Main Theme	4	04:26
334474	Prologue	6	02:12
447670	Duel of the Fates	7	04:14
453112	Opening Titles	8	00:33
453118	Welcome to Jurassic Park	8	07:54
453119	My Friend, the Brachiosaurus	6	04:16
464602	Ah, Rats!!!	8	03:41
464603	Escape From Venice	6	04:24
464604	No Ticket	6	02:46
496934	E.T.'s New Home	4	01:38
515742	Yoda's Theme	6	03:29
515746	Yoda and the Force	6	04:02
515748	Lando's Palace	6	03:53
532750	Star Wars Main Title / Ambush on Corus…	8	03:46
532751	Across the Stars (Love Theme From Atta…	8	05:33
532752	Zam the Assassin / The Chase Through C…	10	11:07
532754	Departing Coruscant	6	01:44
532755	Anakin and Padmé	6	03:56
532757	The Meadow Picnic	8	04:14
532758	Bounty Hunter's Pursuit	8	03:23
532759	Return to Tatooine	8	06:56
532760	The Tusken Camp / The Homestead	6	05:54
532761	Love Pledge / The Arena	8	08:29
532762	Confrontation With Count Dooku / Finale	8	10:45

This table requires foreign key constraints:
e.g, there should be no a.id without a matching artist.id!

Example Data Models: Relational Model

Pros:

- Very expressive
- Not focuses on a specific query, very flexible,
- Relational DBMS can perform queries very fast
  - …and they have other advantages, too!

Cons:

- Requires a DBMS to manage tables, schemas,constraints, and queries
- Complex to use
- Complex to design

Example Data Models: Single-Table Model

- **Data Model**: “A formal definition of how to represent data (in general) and the available data operations.”
- Common Data Model 2: **Single Table Model**
  - Represents all data as a single table
  - Typically: This table can be easily represented by a CSV or EXCEL file
    - No DBMS necessary to manage those files (there are however DBMS specialized in this)
    - File can be / must be directly manipulated by humans or software
    - … but the lack of standardized DBMS makes querying and handling hard
  - Data files using this model files are mostly just used for exchanging data

Example: A logical schema of our music database in the single table model

- We need to choose which aggregation perspective to adapt!

Option 1?

ArtistRecording
artist_id
artist_name
recording_id
recording_name
rating
length

Option 2?

ArtistAlbum
artist_id
artist_name
album_id
album_name
album_release_date

Option 3?

ArtistAlbumRecording
artist_id
artist_name
album_id
album_name
album_release_date
recording_album_pos
recording_id
recording_rating
recording_ength

Example: A logical schema of our music database in the single table model

- We need to choose which aggregation perspective to adapt!

Option 1?

ArtistRecording
artist_id
artist_name
recording_id
recording_name
rating
length

Option 2?

ArtistAlbum
artist_id
artist_name
album_id
album_name
album_release_date

Option 3?

ArtistAlbumRecording
artist_id
artist_name
album_id
album_name
album_release_date
recording_album_pos
recording_id
recording_rating
recording_ength

To clarify: The question here is is how to aggregate a conceptual schema with 3 entities and 3 relationships into a single table..:!

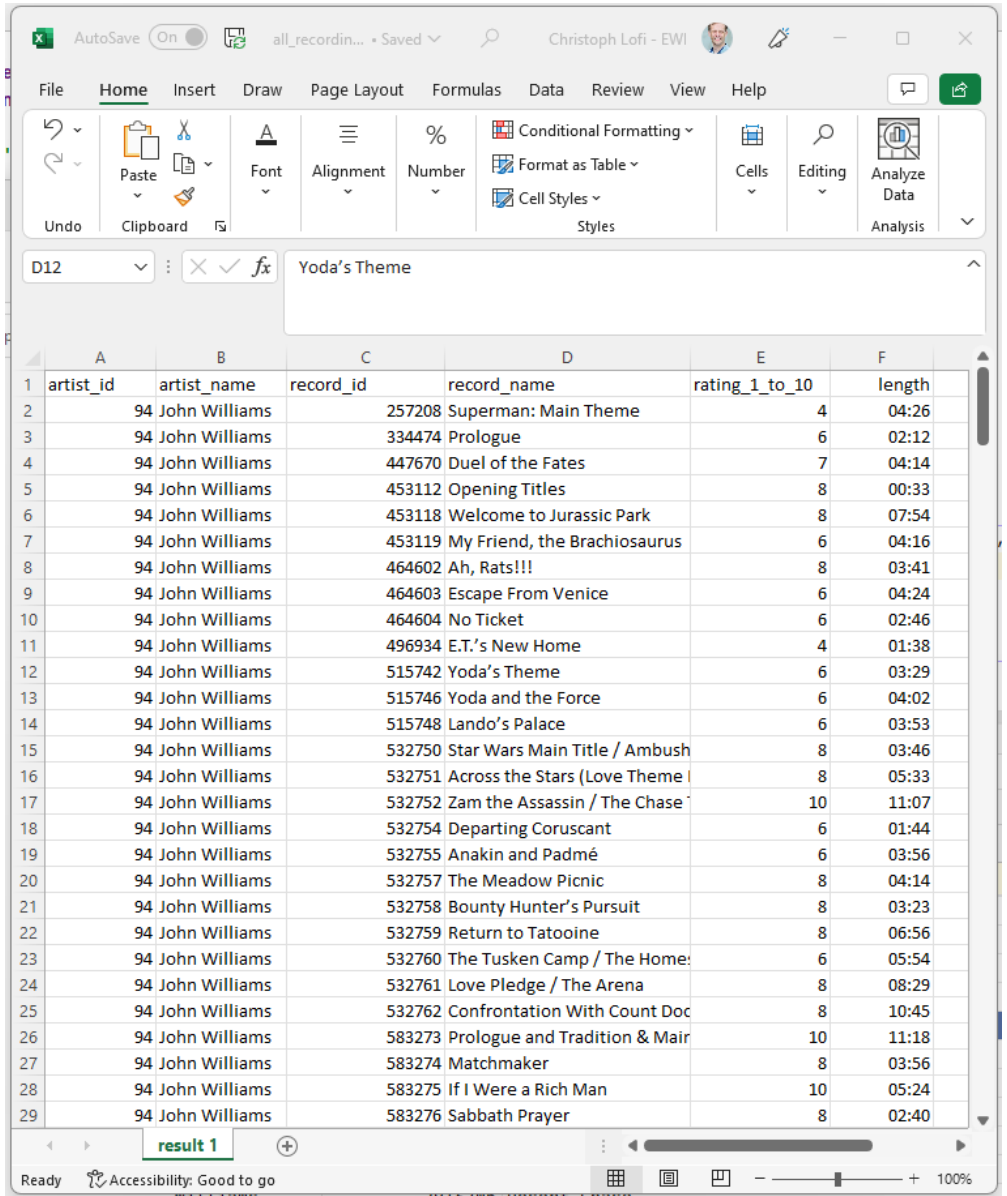
Example: A logical schema of our music database in the single table model

- We need to choose which aggregation perspective to adapt

Option 1?

ArtistRecording
artist_id
artist_name
recording_id
recording_name
rating
length

as **Excel**



	A	B	C	D	E	F
1	artist_id	artist_name	record_id	record_name	rating_1_to_10	length
2	94	John Williams	257208	Superman: Main Theme	4	04:26
3	94	John Williams	334474	Prologue	6	02:12
4	94	John Williams	447670	Duel of the Fates	7	04:14
5	94	John Williams	453112	Opening Titles	8	00:33
6	94	John Williams	453118	Welcome to Jurassic Park	8	07:54
7	94	John Williams	453119	My Friend, the Brachiosaurus	6	04:16
8	94	John Williams	464602	Ah, Rats!!!	8	03:41
9	94	John Williams	464603	Escape From Venice	6	04:24
10	94	John Williams	464604	No Ticket	6	02:46
11	94	John Williams	496934	E.T.'s New Home	4	01:38
12	94	John Williams	515742	Yoda's Theme	6	03:29
13	94	John Williams	515746	Yoda and the Force	6	04:02
14	94	John Williams	515748	Lando's Palace	6	03:53
15	94	John Williams	532750	Star Wars Main Title / Ambush	8	03:46
16	94	John Williams	532751	Across the Stars (Love Theme	8	05:33
17	94	John Williams	532752	Zam the Assassin / The Chase	10	11:07
18	94	John Williams	532754	Departing Coruscant	6	01:44
19	94	John Williams	532755	Anakin and Padmé	6	03:56
20	94	John Williams	532757	The Meadow Picnic	8	04:14
21	94	John Williams	532758	Bounty Hunter's Pursuit	8	03:23
22	94	John Williams	532759	Return to Tatooine	8	06:56
23	94	John Williams	532760	The Tusken Camp / The Homes	6	05:54
24	94	John Williams	532761	Love Pledge / The Arena	8	08:29
25	94	John Williams	532762	Confrontation With Count Doc	8	10:45
26	94	John Williams	583273	Prologue and Tradition & Mair	10	11:18
27	94	John Williams	583274	Matchmaker	8	03:56
28	94	John Williams	583275	If I Were a Rich Man	10	05:24
29	94	John Williams	583276	Sabbath Prayer	8	02:40

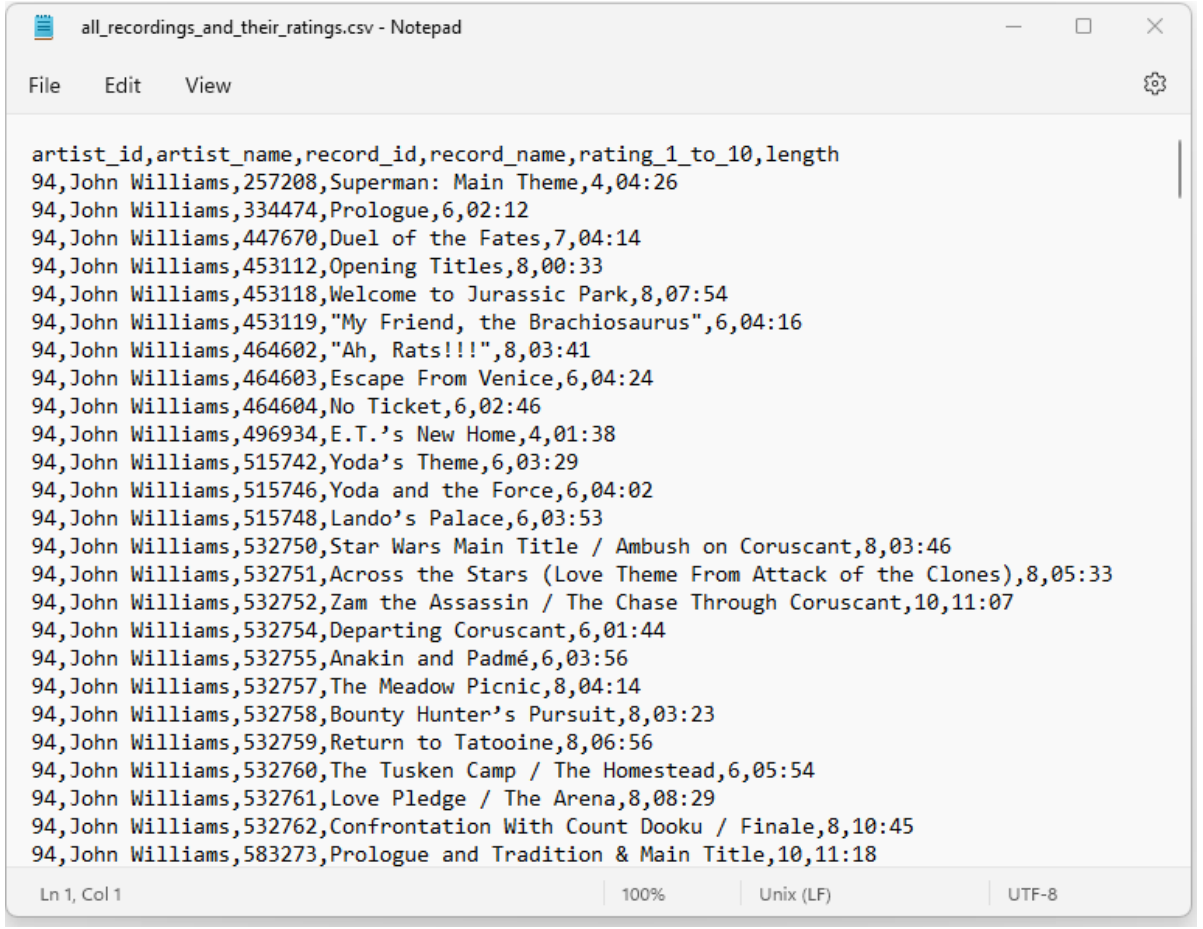
Example: A logical schema of our music database in the single table model

- We need to choose which aggregation perspective to adapt

Option 1?

ArtistRecording
artist_id
artist_name
recording_id
recording_name
rating
length

as **CSV** (comma separated value)



Example Data Models: Single Table Model

Pros:

- Very simple
- Can easily be converted to files
- No DBMS necessary

Cons:

- Inflexible - hard to perform more than one meaningful query
- Schema is “flattened” – most conceptual schema information is lost
- A lot of data redundancy depending on how the schema was flattened

Example Data Models: Document Model

Data Model: "A formal definition on how to represent data (in general) and the available data operations."

Common Data Model 3: **Document Model**

- Represents as semi-structured text documents
- Typically: Decide on a main entity type, and *embed all* its related sub-entities in a substructure
  - Can be easily turned into a file, like JSON or XML
- Queries can be used to filter the document
- Often used in Web APIs and Web systems

Example: A logical schema of our music database in the document model

Option 1?

ArtistRecording

```
Artist {
  id
  name
  Recording {
    id
    name
    rating
    length
  }
}
```

Option 2?

ArtistAlbum

```
Artist {
  id
  name
  Album {
    id
    name
    releasedate
  }
}
```

Option 3?

RecordAlbum

```
Record {
  id
  name
  length
  Album {
    id
    name
    position
  }
}
```

Example: A logical schema of our music database in the document model

Option 1?

ArtistRecording

```
Artist {
  id
  name
  Recording {
    id
    name
    rating
    length
  }
}
```

Option 2?

ArtistAlbum

```
Artist {
  id
  name
  Album {
    id
    name
    releasedate
  }
}
```

Option 3?

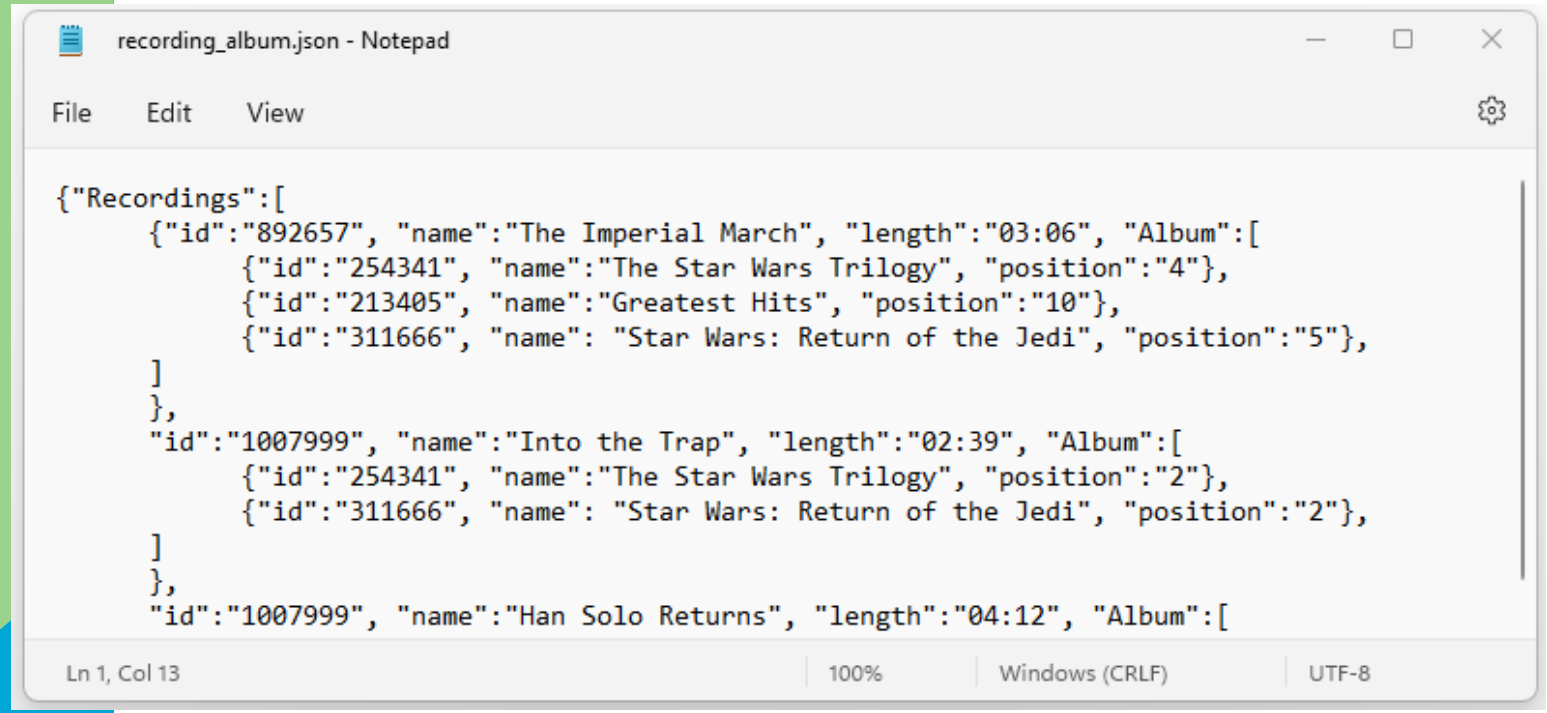
RecordAlbum

```
Record {
  id
  name
  length
  Album {
    id
    name
    position
  }
}
```

To clarify: The question here is is how to aggregate a conceptual schema with 3 entities and 3 relationships into semi-structured document tree!

Example: A logical schema of our music database in the document model

- We need to choose (again) an aggregation perspective



Option 3?

RecordAlbum

```
Record {
  id
  name
  length
  Album {
    id
    name
    position
  }
}
```

as JSON (JavaScript Object Notation)

Example Data Models: Document Model

Pros:

- Simple
- Can easily be converted to files
- Some structure remains (compared to single table model)
- No DBMS needed (but there are good Document DBMS)
- For later: very scalable and distributable!

Cons:

- Somewhat inflexible – Queries can be performed but they are typically less powerful than for relational model
- Some data redundancy depending on the aggregation perspective

Summary: What did we learn?

- Databases are collection of related data
- Databases adopt a data model, and are structured using a schema
  - Conceptual Schema: Describes entities and their relationships
  - Logical Schema: Represents the Conceptual Schema in the chosen Data Model
- The **data model** governs how data can be represented and manipulated

Relational Model	Single Table Model	Document Model	For Later: Graph Model
• Uses tables with attributes; tables can be “linked” via Foreign Key Constraints • Implemented by traditional database management systems like PostgreSQL, Oracle, MySQL, etc.	• Just one table typically without additional constraints • CSV or EXCEL files	• Uses a semi-structured document tree • JSON or XML files, but also non-relational DBMS like CouchDB or MongoDB	• Uses a graph with nodes and edges • RDF or GraphDBs like Neo4j

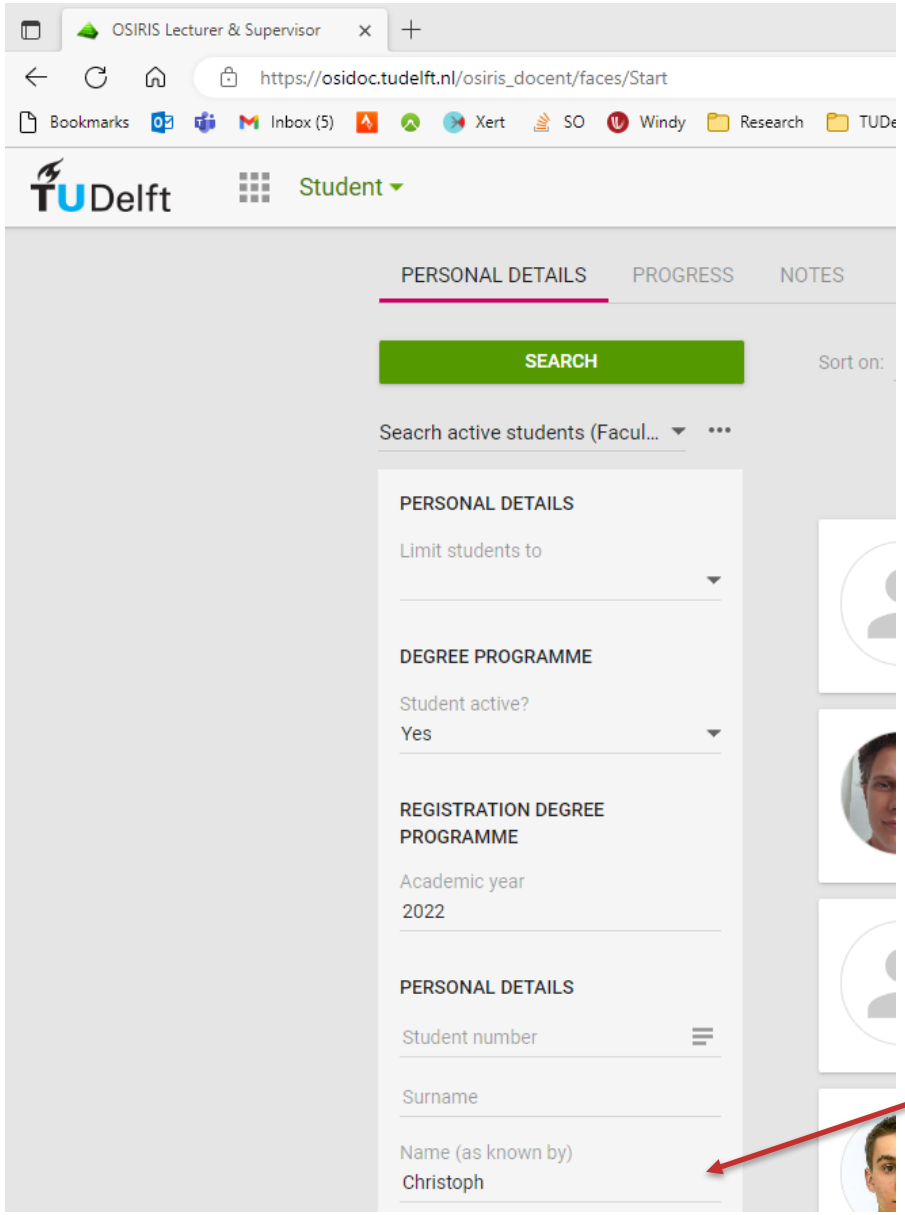
Why should you care about IDM?

The Role of Relational Databases

The Role of Relational Databases

The Role of Relational Databases

???



The exam room list takes 6 minutes to generate???

7 Seconds for a simple search-by-name?

Information and Data Management