Datenmodellierung/Datenbanksysteme

VU 184.685/VU 184.686, WS 2022

Einführung

Johannes Fichte, Felix Winter

Institut für Logic and Computation
Technische Universität Wien

Acknowledgments

Die Folien sind eine Weiterentwicklung der Folien von Katrin Seyr und Sebastian Skritek.

Der Inhalt basiert auf und behandelt Kapitel 1 des Lehrbuchs
(Kemper, Eickler: Datenbanksysteme – Eine Einführung)

Was haben diese Firmen gemeinsam?

Motivation

Wir haben heute

- Groï¿½e Datenmengen ...
- ...in den unterschiedlichsten Bereichen
- Verteilte Applikationen
- Kritische Anwendungen

$\Rightarrow$ benötigen Software zur effizienten Verarbeitung

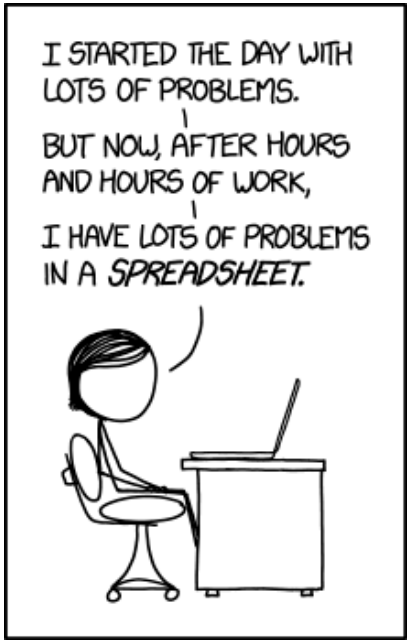
Zusätzliche Anforderungen an die Software

Zu lösende Probleme:

- Redundanz und Inkonsistenz
- Verlust von Daten
- Mehrbenutzerbetrieb
- Sicherheitsprobleme
- Integritätsverletzungen
- Beschränkte Zugriffsmöglichkeiten

Führen zu:

- Komplexer Software
- Hohe Entwicklungskosten



Quelle: https://xkcd.com/1906/

Datenbankverwaltungssystem (DBMS)

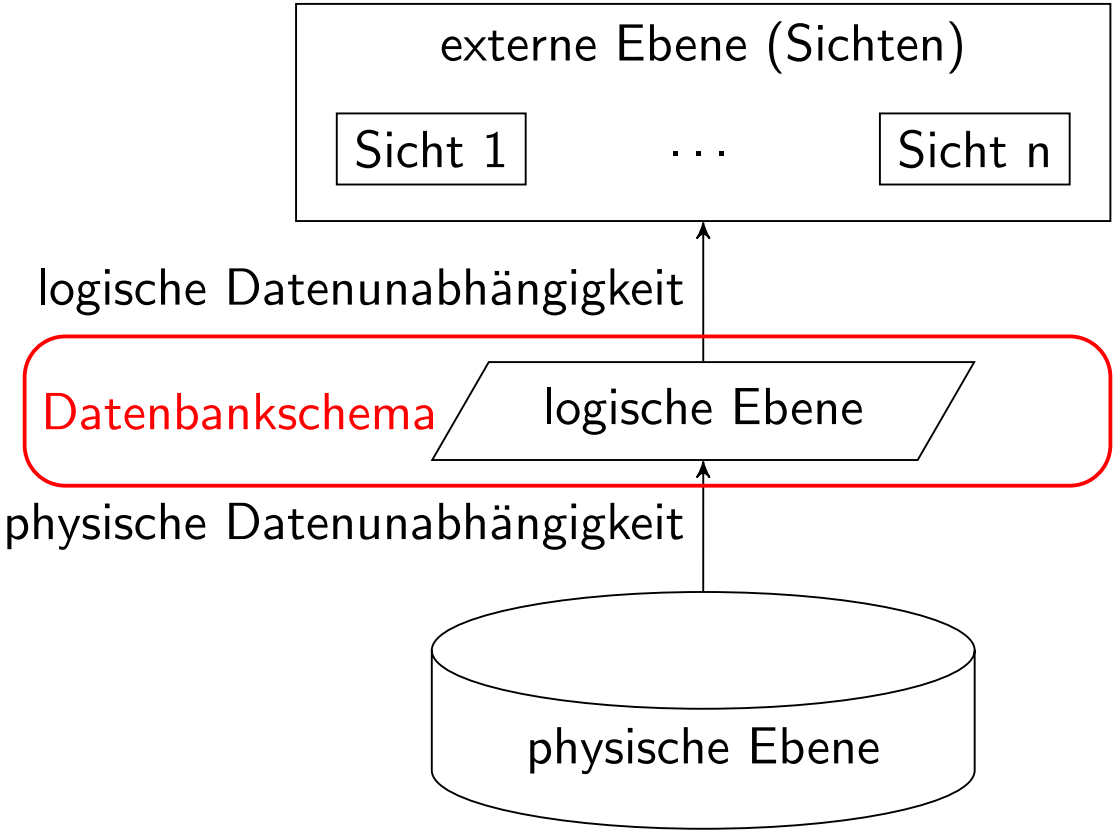
Definition (Datenbankmanagementsystem (DBMS))

Der Begriff Datenbankmanagmentsystem (DBMS) (auch Datenbankverwaltungssystem, engl. *database management system*) bezeichnet die Gesamtheit der Programme zum Zugriff auf die (im DBMS) gespeicherten Daten.

Definition (Datenbasis)

Als Datenbasis bezeichnet man die in einem DBMS gespeicherten Daten.

Datenabstraktion und Datenunabhängigkeit



Datenbankschema- und Ausprägung

Definition (Datenbankschema)

Das Datenbankschema legt die Struktur der Daten fest. (Metadaten: Daten über die Datenbasis.)

Definition (Datenbankausprägung)

Eine Datenbankausprägung ist eine konkrete Datenbasis welche jener im Datenbankschema beschriebenen Struktur folgt und sämtlichen fort festgelegten Bedingungen erfüllt.

Datenmodell

Intuitiv: Definiert die Sprache und Werkzeuge zur Beschreibung des Datenbankschemas und zur Manipulation/Interaktion mit Datenbankausprägungen.

Datenmodelle:

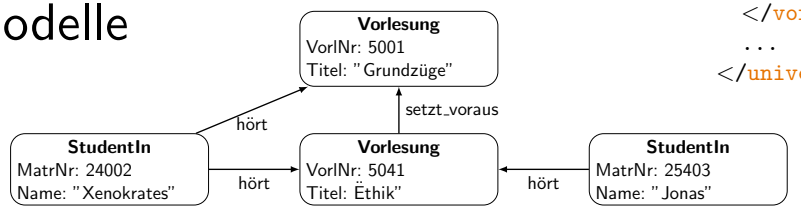
- Netzwerkmodell
- Hierarchisches Modell
- Relationales Modell
- Objekt Orientierte Datenmodelle
- XML
- Graph-Datenmodelle
- ...

StudentIn	
MatrNr	Name
24002	Xenokrates
25403	Jonas
...	...

hï¿½ren	
MatrNr	VorlNr
26120	5001
...	...

Vorlesung	
VorlNr	Titel
5001	Grundzï¿½ge
5041	Ethik
...	...

```xml
<?xml version="1.0">
<university>
  <vorlesung vorlnr="5001" titel="Grundzüge">
    <gehoert_von>
      <student matrnr="24002" name="xenokrates/>
      <student matrnr="25403" name="jonas/>
      ...
    </gehoert_von>
    ...
  </vorlesung>
  ...
</university>
```



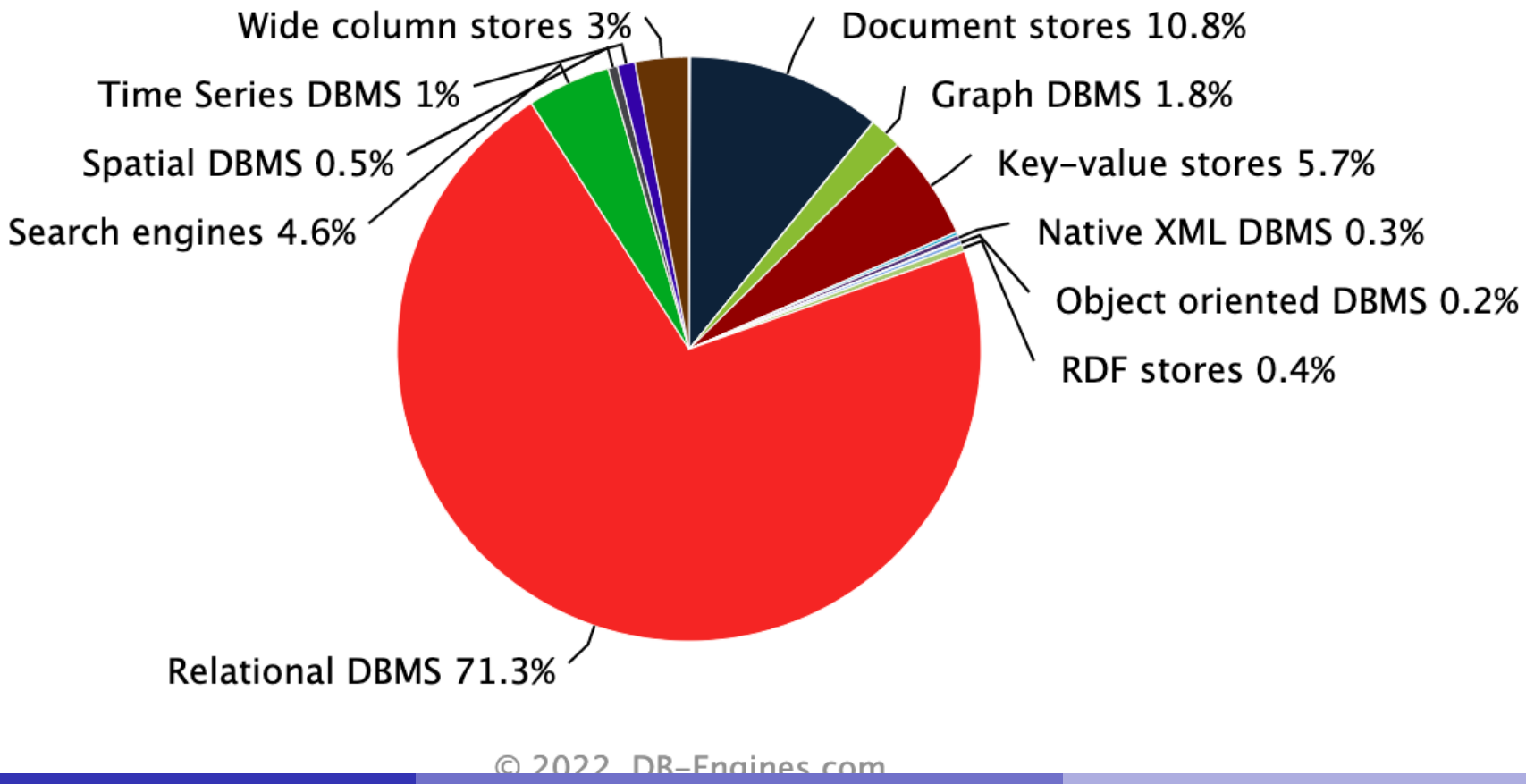
Datenmodell

Definition (Datenmodell)

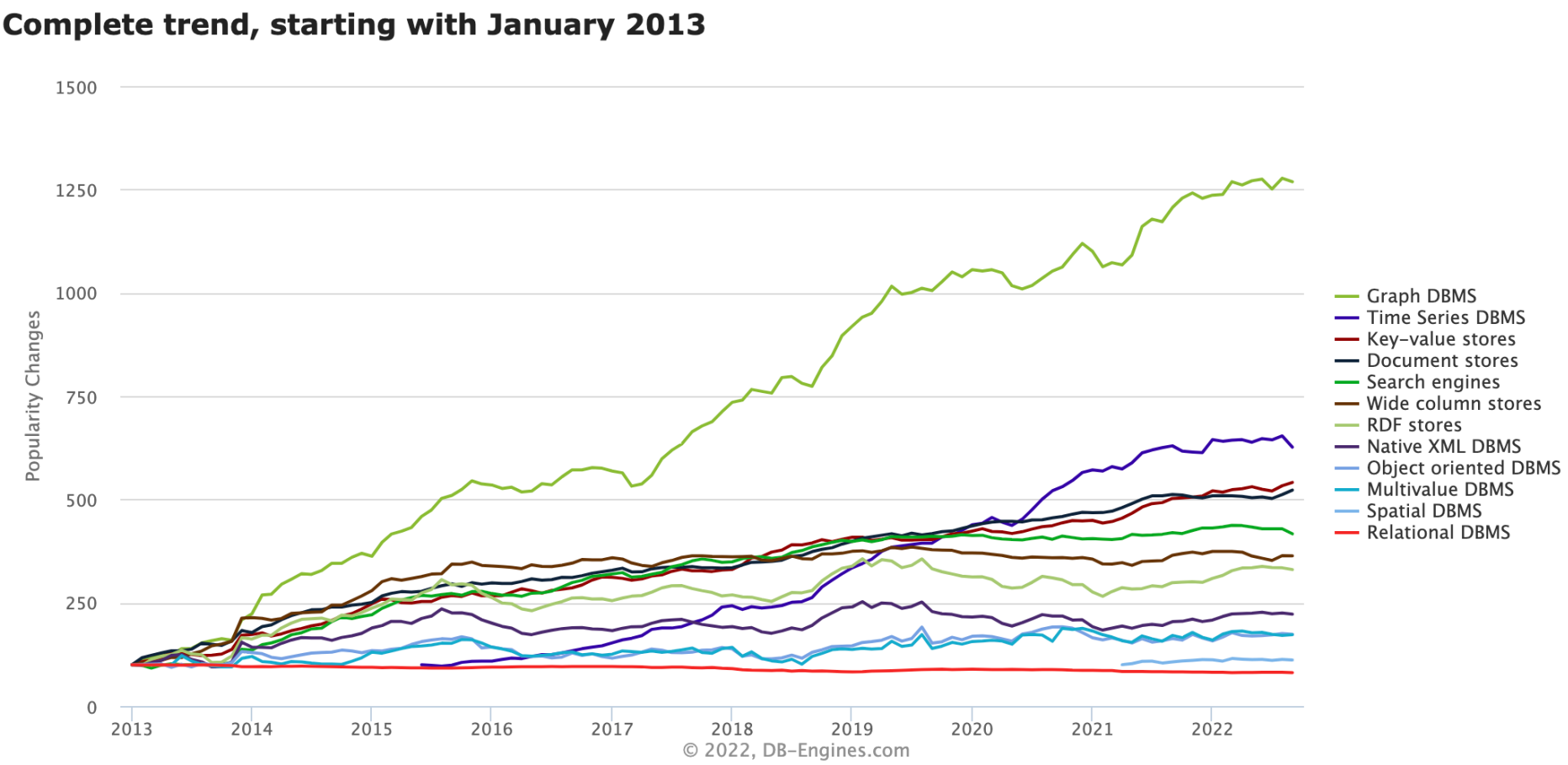
Ein Datenmodell legt die anwendbaren Strukturen und Operatoren fest. Es besteht aus:

- Einer Datendefinitionssprache (DDL) (*data definition language*) zur Schemadefinition
- Einer Datenmanipulationssprache (DML) (*data manipulation language*) zur Interaktion mit und Manipulation von der Ausprägung. Man unterscheidet meist die
  - Abfragesprache (auch Anfragesprach; *query language*)
  - der eigentlichen Datenmanipulationssprache (Befehlen zum Einfügen, Löschen, ...)

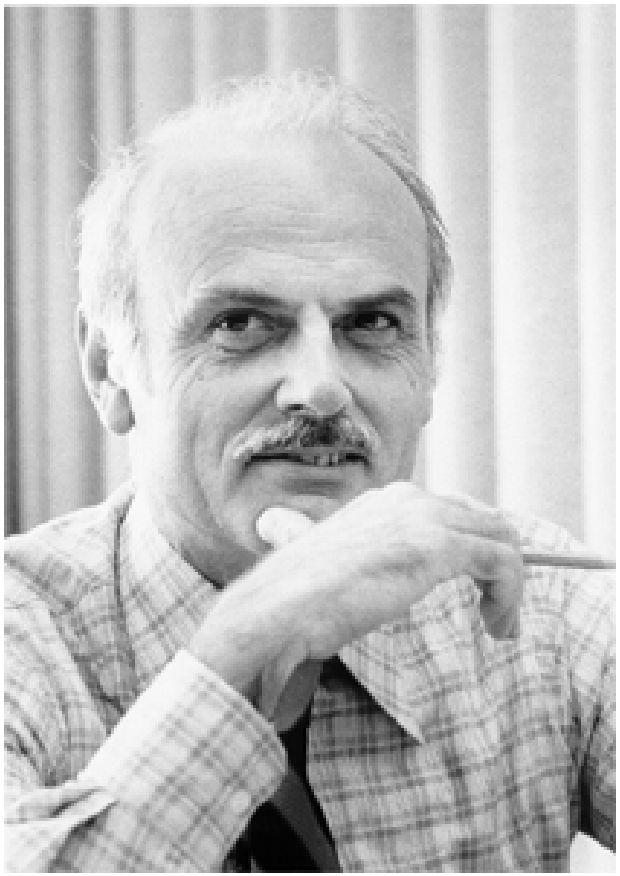
Datenmodelle



Datenmodelle



Das Relationale Datenmodell



Edgar F. Codd (1923-2003)

A Relational Model of Data for Large Shared Data Banks in: Communications of the ACM, Volume 13, Issue 6 (June 1970) Pages: 377 - 387 (1970)

Turing Award (1981) $\approx$ Nobelpreis der Informatik

Das Relationale Datenmodell

Anforderungen an das Modell nach E.F. Codd

- Datenintegration (einheitliche Verwaltung der Daten)
- Operationen (Speichern, Suchen, ï¿½ndern)
- Katalog (Zugriff auf Datenbeschreibung)
- Benutzersichten (applikationsspezifische Sicht)
- Konsistenzï¿½berwachung (logische Korrektheit der Daten)
- Zugriffskontrolle (Datenschutz)
- Transaktionen (unteilbare Operationsblï¿½cke)
- Synchronisation (Mehrbenutzerbetrieb)
- Datensicherung (Wiederherstellung nach Systemcrash)

Moderne Relationale DBMS

- DDL/DML: SQL (Structured Query Language)
- Einbettung in Programmiersprachen
- Tools (zB Entwurf, Masken, Interaktion, ...)
- Mehrbenutzerbetrieb, Sicherheitsmechanismen

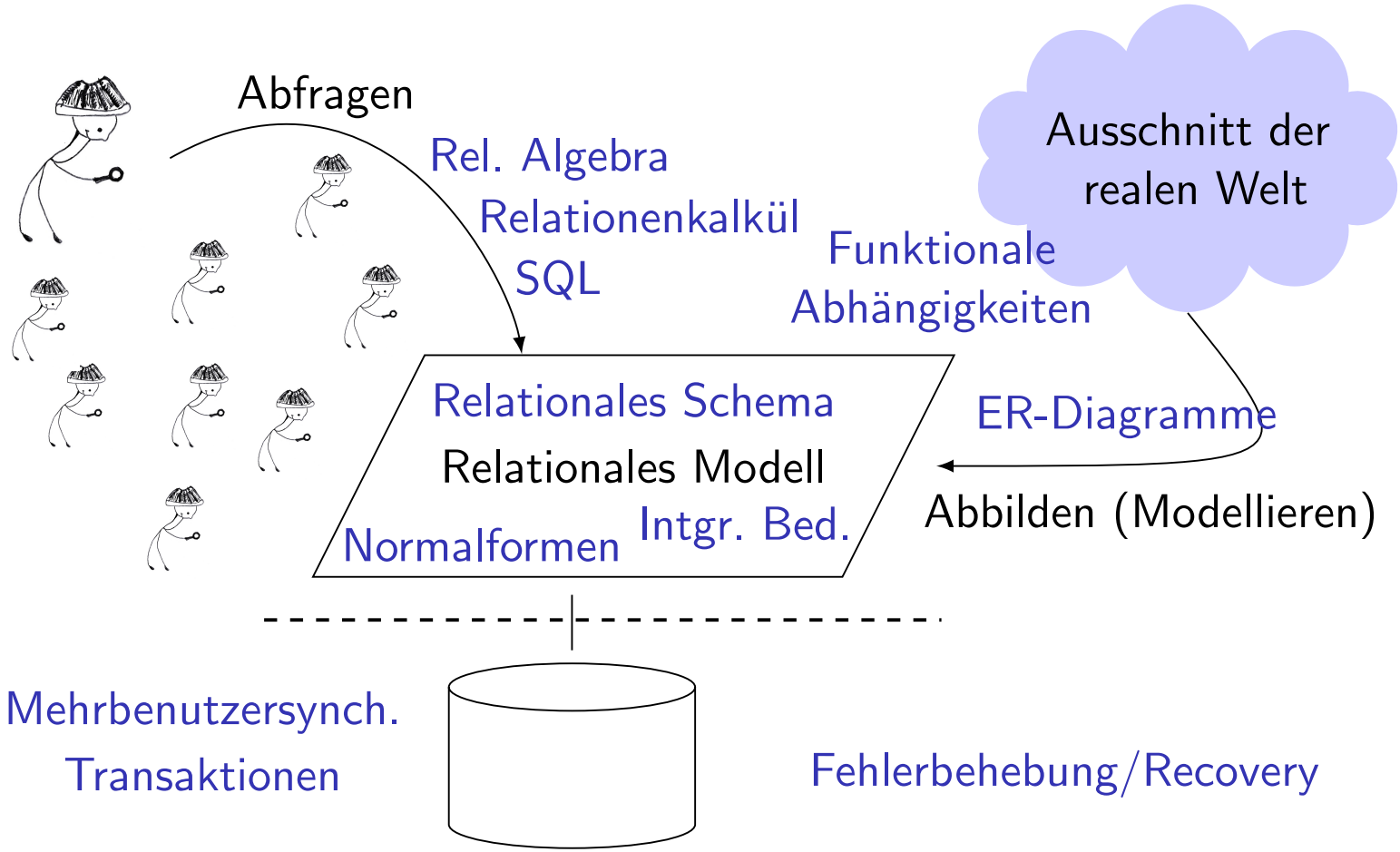
Oracle, IBM DB2, MS SQL Server, PostgreSQL, MySQL, IBM Informix, SAP Sybase, SAP Hana, Teradata, SQLite, . . .

395 systems in ranking, September 2022

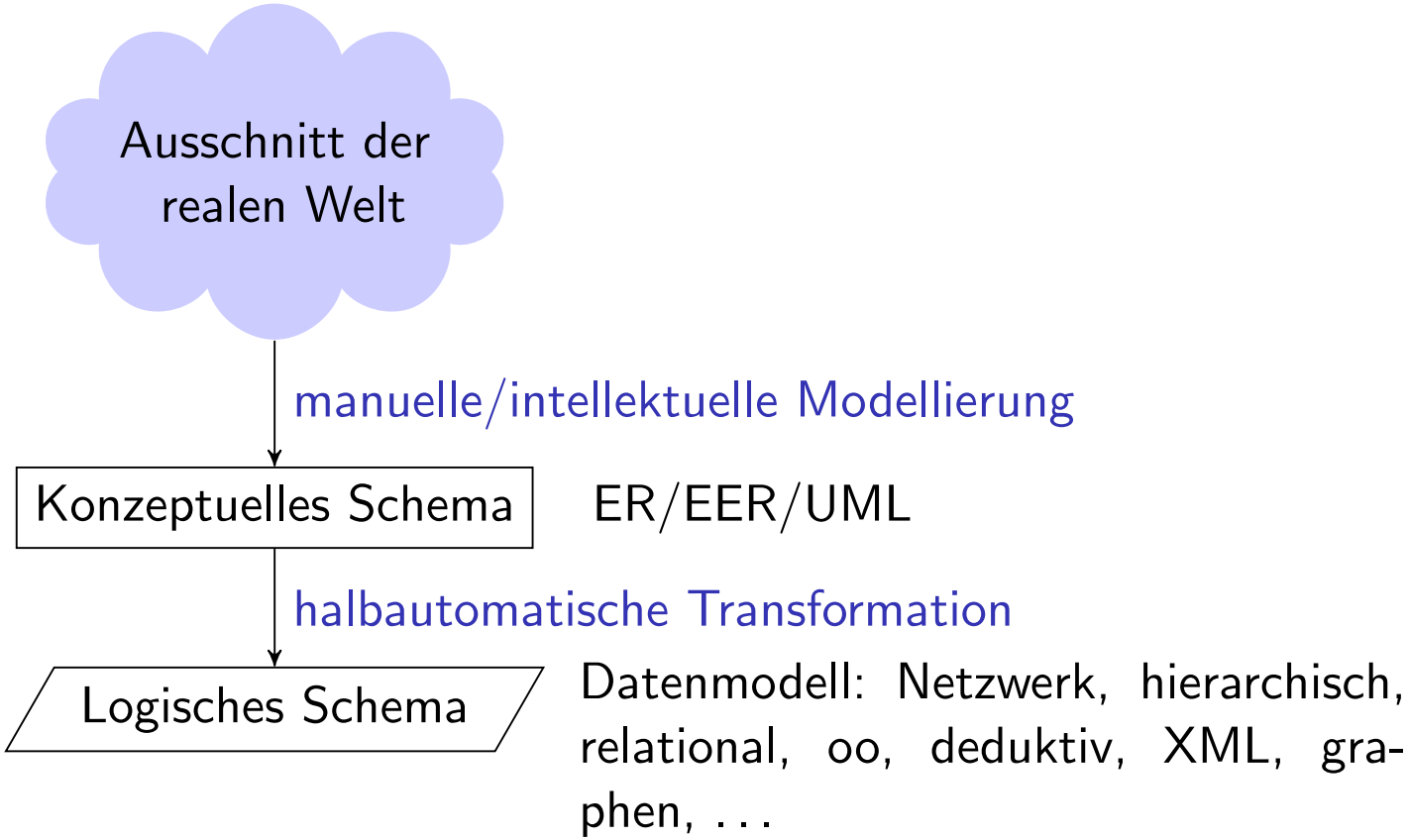
Rank			DBMS	Database Model	Score		
Sep 2022	Aug 2022	Sep 2021			Sep 2022	Aug 2022	Sep 2021
1.	1.	1.	Oracle	Relational, Multi-model	1238.25	-22.54	-33.29
2.	2.	2.	MySQL	Relational, Multi-model	1212.47	+9.61	-0.06
3.	3.	3.	Microsoft SQL Server	Relational, Multi-model	926.30	-18.66	-44.55
4.	4.	4.	PostgreSQL	Relational, Multi-model	620.46	+2.46	+42.95
5.	5.	5.	MongoDB	Document, Multi-model	489.64	+11.97	-6.87
6.	6.	6.	Redis	Key-value, Multi-model	181.47	+5.08	+9.53
7.	↑ 8.	↑ 8.	Elasticsearch	Search engine, Multi-model	151.44	-3.64	-8.80
8.	↓ 7.	↓ 7.	IBM Db2	Relational, Multi-model	151.39	-5.83	-15.16
9.	9.	↑ 11.	Microsoft Access	Relational	140.03	-6.47	+23.09
10.	10.	↓ 9.	SQLite	Relational	138.82	-0.05	+10.17

Quelle: `https://db-engines.com/en/ranking`

Inhalt der LVAs

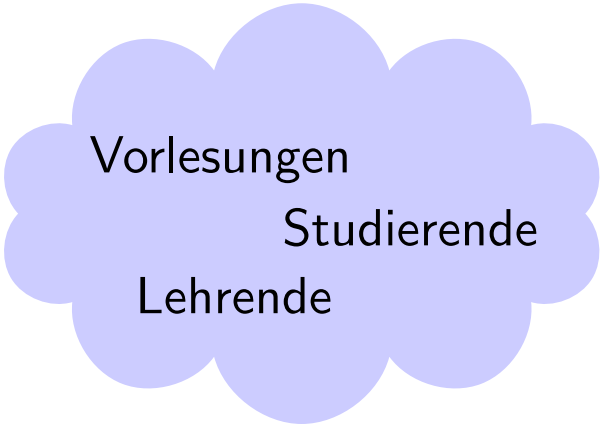


Datenmodellierung



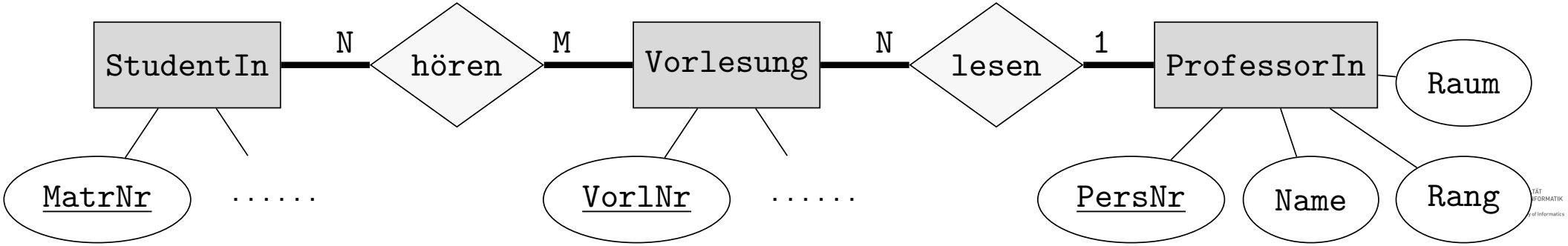
Datenmodellierung (Bsp.)

1. Abgrenzung der zu modellierenden Welt



reale Welt: Universität

2. Überführung der zu modellierenden Welt in ein konzeptuelles Schema (in der VL: EER)



Datenmodellierung (Bsp.)

3. Überführung des konzeptuellen Schemas in ein logisches Schema (in der VL: Relationales Schema)

StudentIn

MatrNr	Name
24002	Xenokrates
25403	Jonas
26120	Fichte
26830	Aristoxenos
28106	Carnap
29555	Feuerbach
...	...

hï¿½ren

MatrNr	VorlNr
26120	5001
24002	5001
24002	4052
...	...

Vorlesung

VorlNr	Titel
5001	Grundzï¿½ge
5041	Ethik
5049	Mï¿½eutik
4052	Logik
5216	Bioethik
...	...