

# **Datenbanken**

Einführung

---

**Seminaristischer Unterricht**

Prof. Dr.-Ing. Hendrik Gärtner

# Gliederung

- Datenbanksysteme: Eine kleine Einführung
  - Was sind Datenbanken?
  - Wo werden Datenbanken eingesetzt?
  - Warum werden Datenbanken eingesetzt?
- Datenbankentwurfsprozess/Einordnung der Themen der Vorlesung

# Was sind Datenbanken?

Wortstämme:

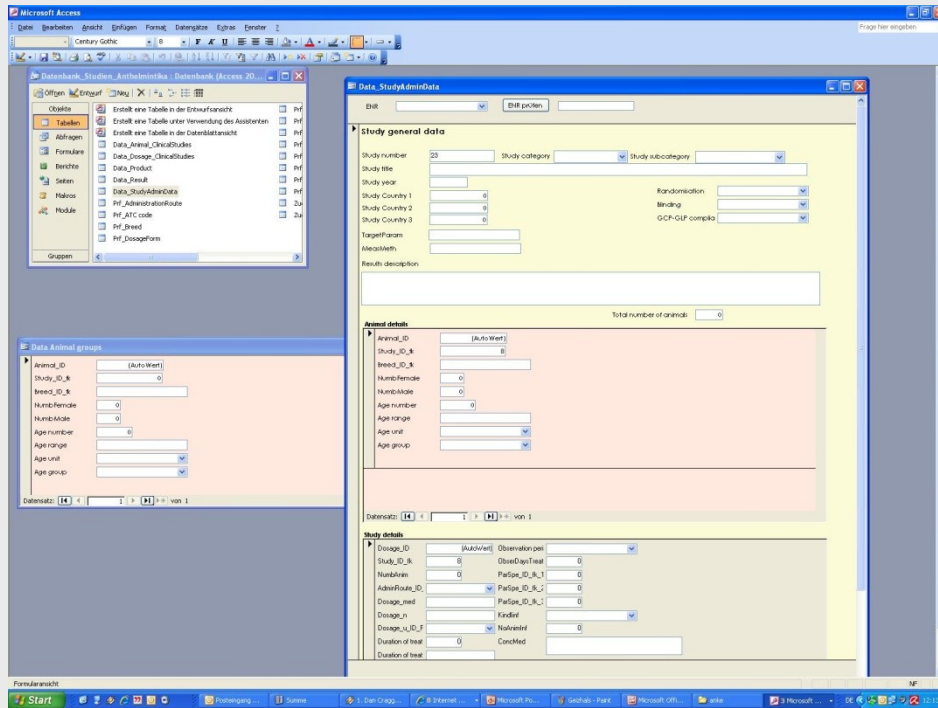
- **Daten:** Logisch strukturierte Informationseinheiten
- **Bank:** Kreditinstitut, das entgeltliche Dienstleistungen für den Zahlungs- und Kredit- und Kapitalverkehr anbietet.

**Analogie zur Bank trifft in vielen Punkten:**

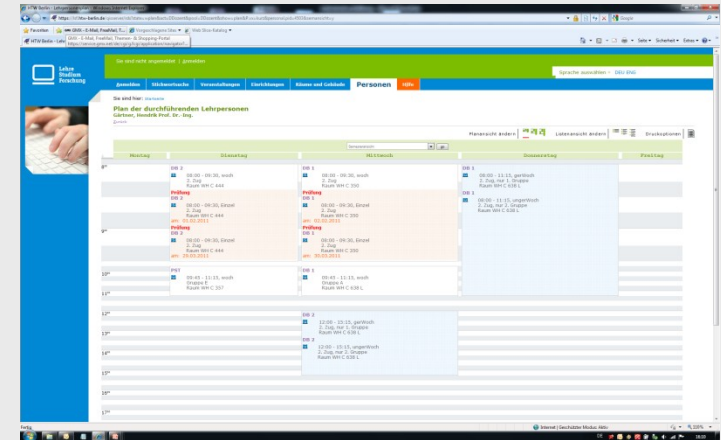
- Die **Sicherheit vor Verlusten** ist eine Hauptmotivation, etwas „auf die Bank zu bringen“.
- Eine Bank bietet **Dienstleistungen für mehrere Kunden** an, um effizient arbeiten zu können.

Eine Datenbank hat die **(langfristige) Aufbewahrung** von Daten als Aufgabe.

# Beispielanwendungen von Datenbanken



Direkter Einsatz von Datenbanken  
in einer klassischen Access-Anwendung



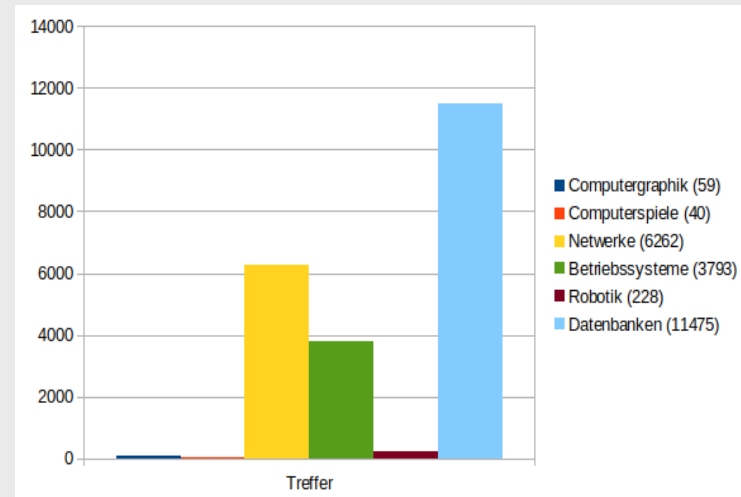
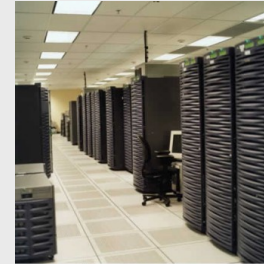
Lehrveranstaltungssystem LSF



Youtube mit Video- und Audiostreaming

# Stellung von Datenbanken in der Informationstechnik

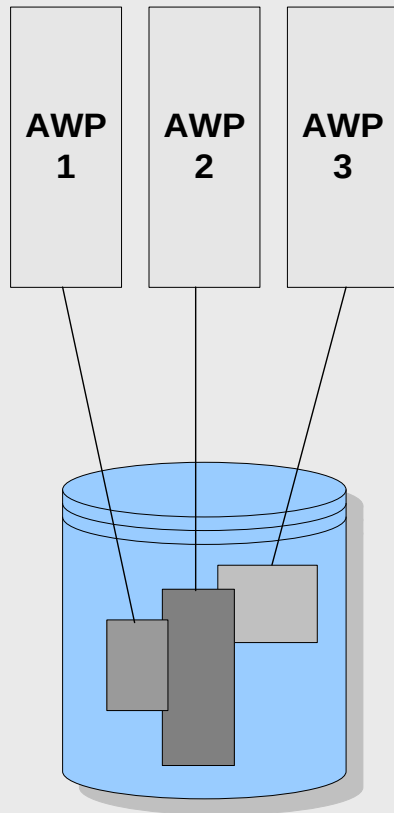
- Datenbanken sind das Herzstück heutiger IT-Infrastrukturen
- ... sind allgegenwärtig
- Datenbankkenntnisse sind stark gefragt



Quelle: Monster 2020

Warum werden Datenbanken eingesetzt und nicht einfach alles auf der Festplatte in Files abgelegt?

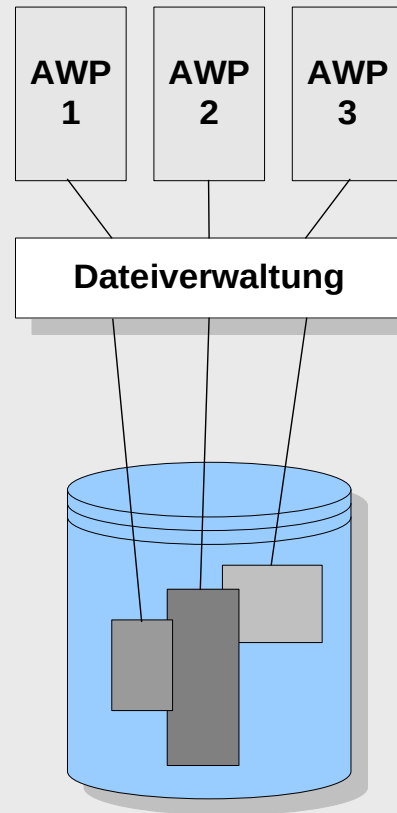
# Historische Entwicklung der Datenverwaltung



Einzel-  
Dateiverwaltung

AWP: Anwenderprogramm

**60'ger Jahre**



Gemeinsame  
Dateiverwaltung

**70'ger Jahre**

## Merkmale:

- Jedes Anwendungssystem verwaltet seine eigenen Daten
- Daten sind mehrfach gespeichert redundant

## Probleme:

- Verschwendung von Speicherplatz
- „Vergessen“ von Änderungen
- Keine zentrale, „genormte“ Datenhaltung

## Weitere Problemstellungen der Ansätze

- Nicht alle Softwaresysteme können große Mengen von Daten effizient verarbeiten
- Mehrere Benutzer oder Anwendungen können nicht parallel auf den gleichen Daten arbeiten (Mehrbenutzerbetrieb)
- Anwendungsprogrammierer / Benutzer können Anwendungen nicht entwickeln / benutzen, ohne:
  - die interne Darstellung der Daten sowie
  - Speichermedien oder Rechner zu kennen (Datenunabhängigkeit nicht gewährleistet)
- Datenschutz und Datensicherheit sind meist nicht gewährleistet



# Datenbank nach Wikipedia

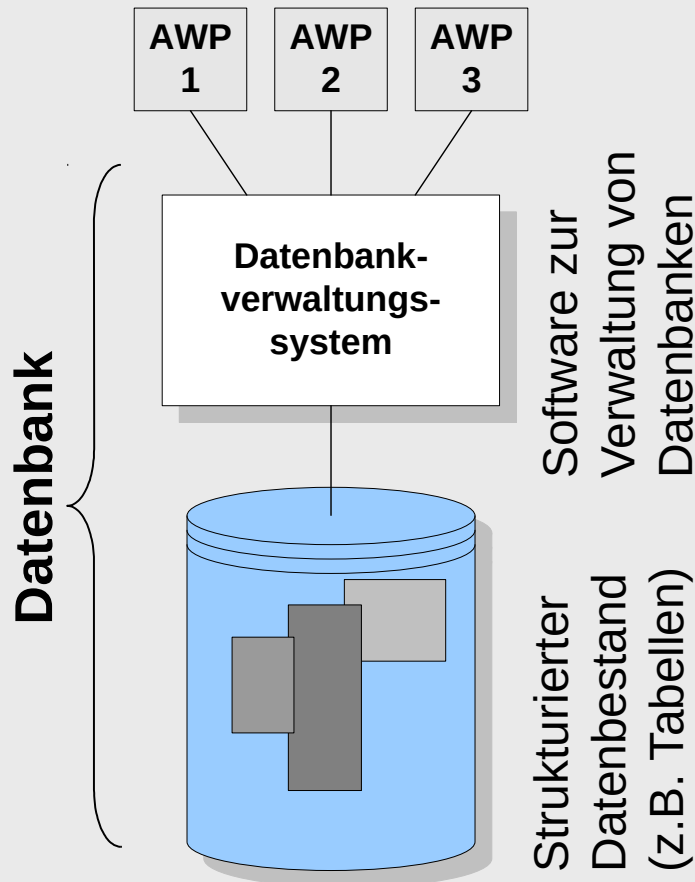
Ein Datenbanksystem (DBS) ist ein System zur **elektronischen Datenverwaltung**. Die wesentliche Aufgabe eines DBS ist es, **große Datenmengen effizient, widerspruchsfrei und dauerhaft zu speichern** und benötigte Teilmengen in unterschiedlichen, bedarfsgerechten Darstellungsformen **für Benutzer und Anwendungsprogramme bereitzustellen**.

Ein DBS besteht aus **zwei Teilen**: der Verwaltungssoftware, genannt **Datenbankmanagementsystem (DBMS)** und der Menge der zu verwaltenden Daten, der **eigentlichen Datenbank (DB)**.

Die Verwaltungssoftware **organisiert intern die strukturierte Speicherung der Daten und kontrolliert alle lesenden und schreibenden Zugriffe** auf die Datenbank.

# Heutige Datenbanken

**Grundidee:** „Genormte Datenhaltung mit einheitlichen Schnittstellen“ für:



- **Reduktion der Entwicklungskosten**
- Vermeidung von Redundanzen,
- Erweiterung der Zugriffsmöglichkeiten (Verknüpfung von Daten),
- Bereitstellung des Mehrbenutzerbetriebs,
- Vermeidung des Datenverlusts,
- Vermeidung von Integritätsverletzungen
- Beschränkungen des Datenzugriffs

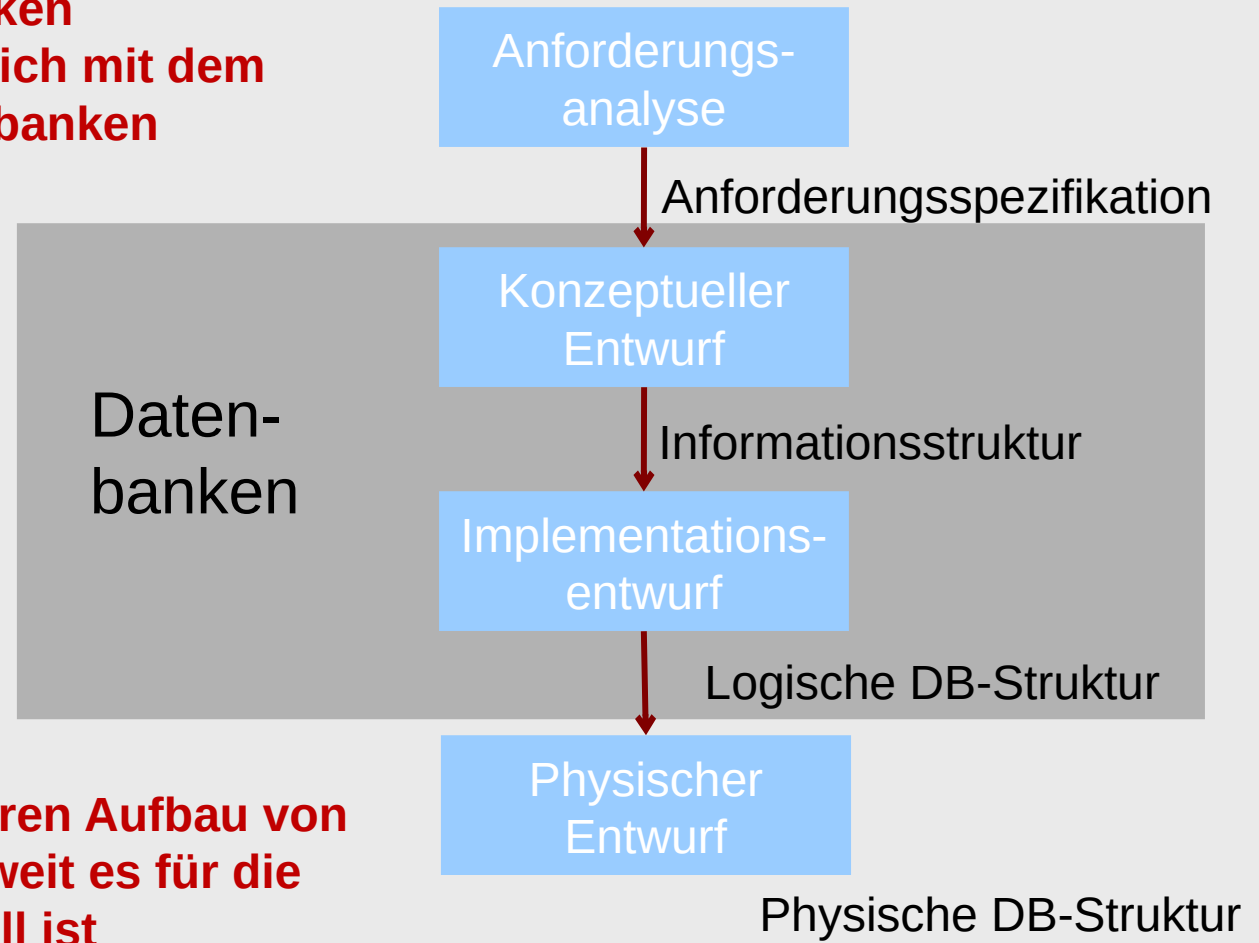
# Prinzipien: Die neun Codd'schen Regeln<sup>1</sup>

1. **Integration:** einheitliche, nichtredundante Datenverwaltung
2. **Operationen:** Speichern, Suchen, Ändern
3. **Katalog:** Zugriffe auf Datenbankbeschreibungen im Data Dictionary
4. **Benutzersichten**
5. **Integritätssicherung:** Korrektheit des Datenbankinhalts
6. **Datenschutz:** Ausschluss unauthorisierter Zugriffe
7. **Transaktionen:** mehrere DB-Operationen als Funktionseinheit
8. **Synchronisation:** parallele Transaktionen koordinieren
9. **Datensicherung:** Wiederherstellung von Daten nach Systemfehlern

<sup>1</sup> Codd'schen Regeln sind für einen zuverlässigen und effizienten Datenbankbetrieb notwendig und hinreichend.

# Phasen des Datenbankentwurfs

**Im Fach Datenbanken beschäftigen Sie sich mit dem Einsatz von Datenbanken**



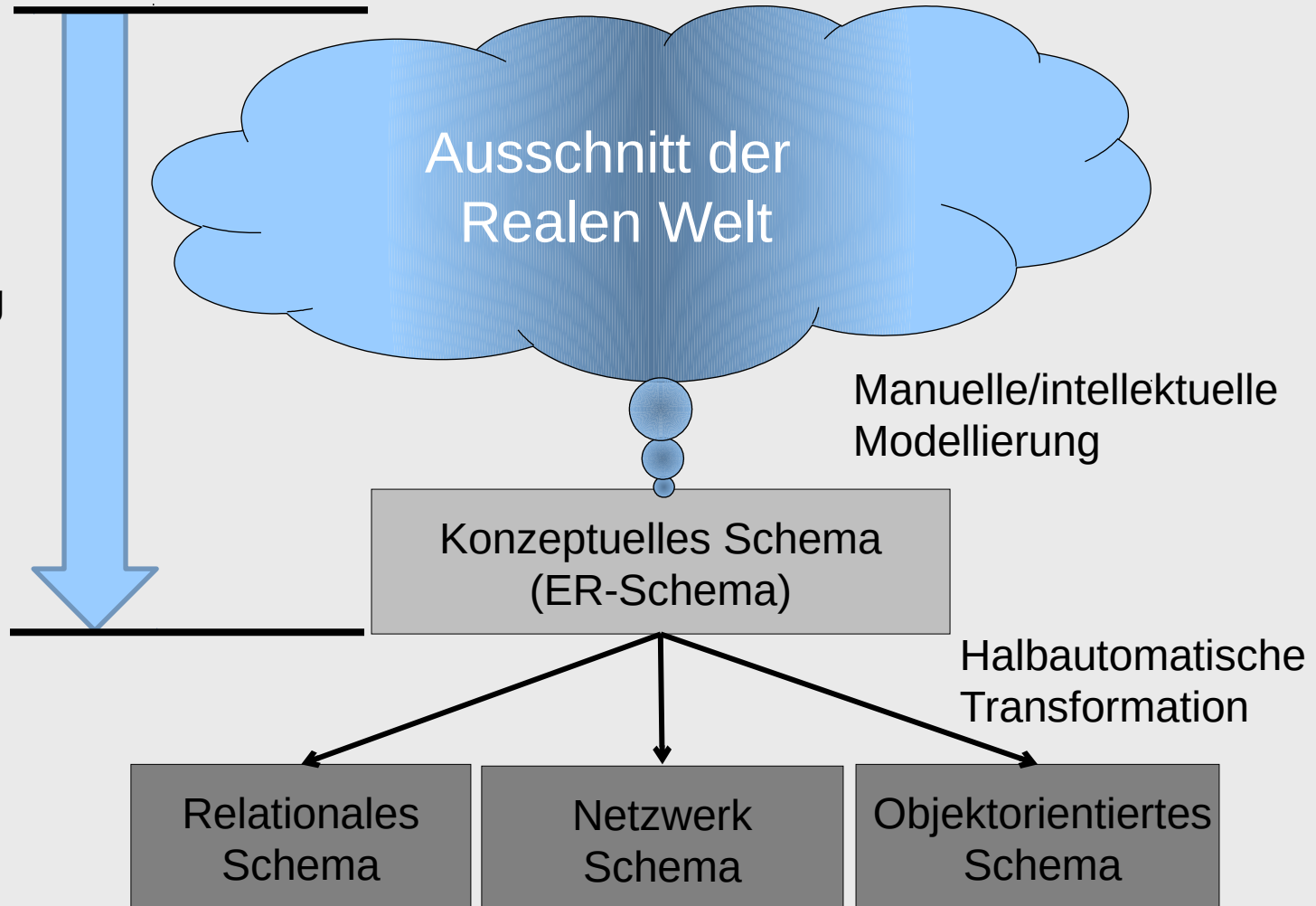
**Sie lernen den inneren Aufbau von DBs – aber nur so weit es für die Anwendung sinnvoll ist**

# Datenbankentwurf: Erster Schritt

## Kreativer Prozess

Schwierigster Teil der Datenbankentwicklung

Abbildung der Wirklichkeit innerhalb eines Modells



# Konzeptuelles Modell

- **Ziel:** Betrieblicher Vorgang soll durch Informationstechnik unterstützt werden
- **Problem:** Chaos der realen Welt muss in ein realitätsnahes vom Computer verarbeitbares Modell umgesetzt werden

Erster Schritt:

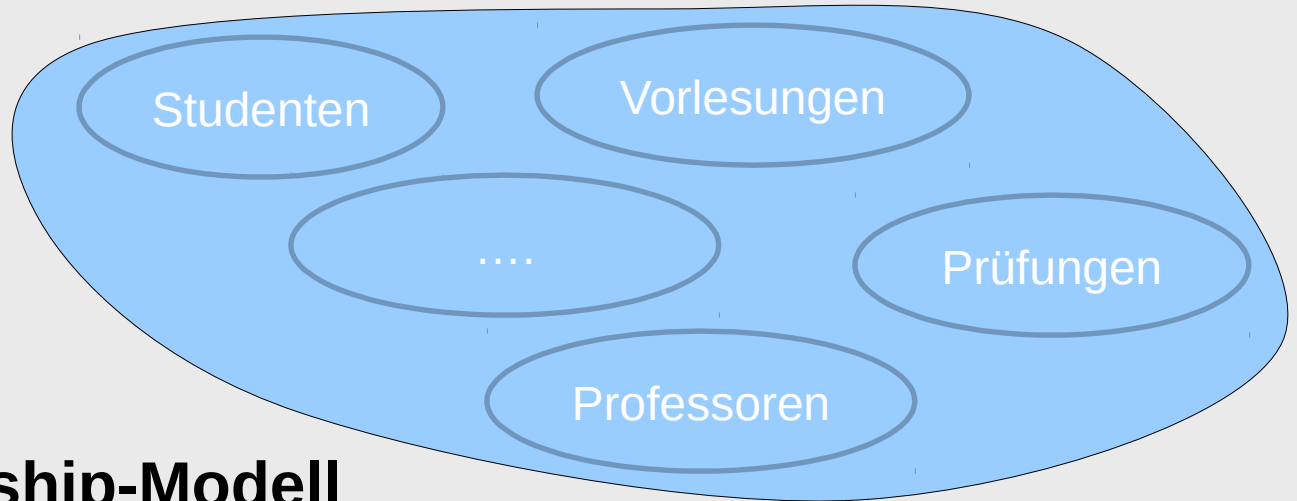
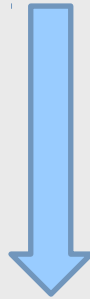
**Abbildung der Situation in Form von Objekten und Beziehungen**

Abbildung erfolgt unabhängig von der zu Grunde liegenden Datenbank, d.h. es wird noch nicht in Tabellen gedacht sondern es findet eine grundlegende Modellierung des Sachverhalts statt!!!!

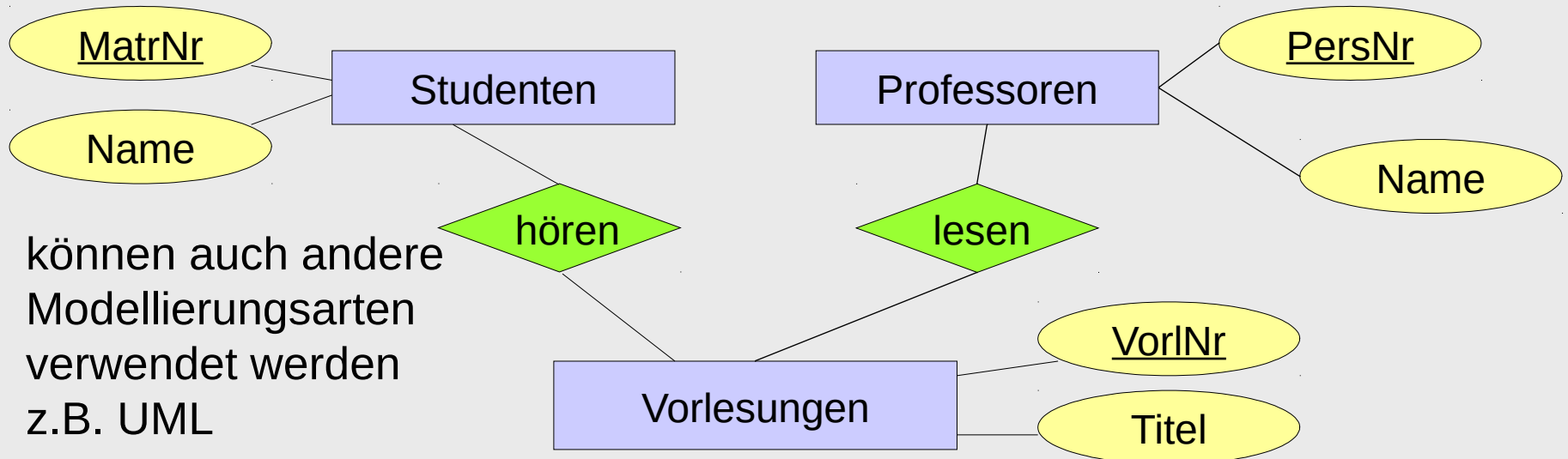
# Beispiel Konzeptuelles Schema

## Mini-Welt HTW

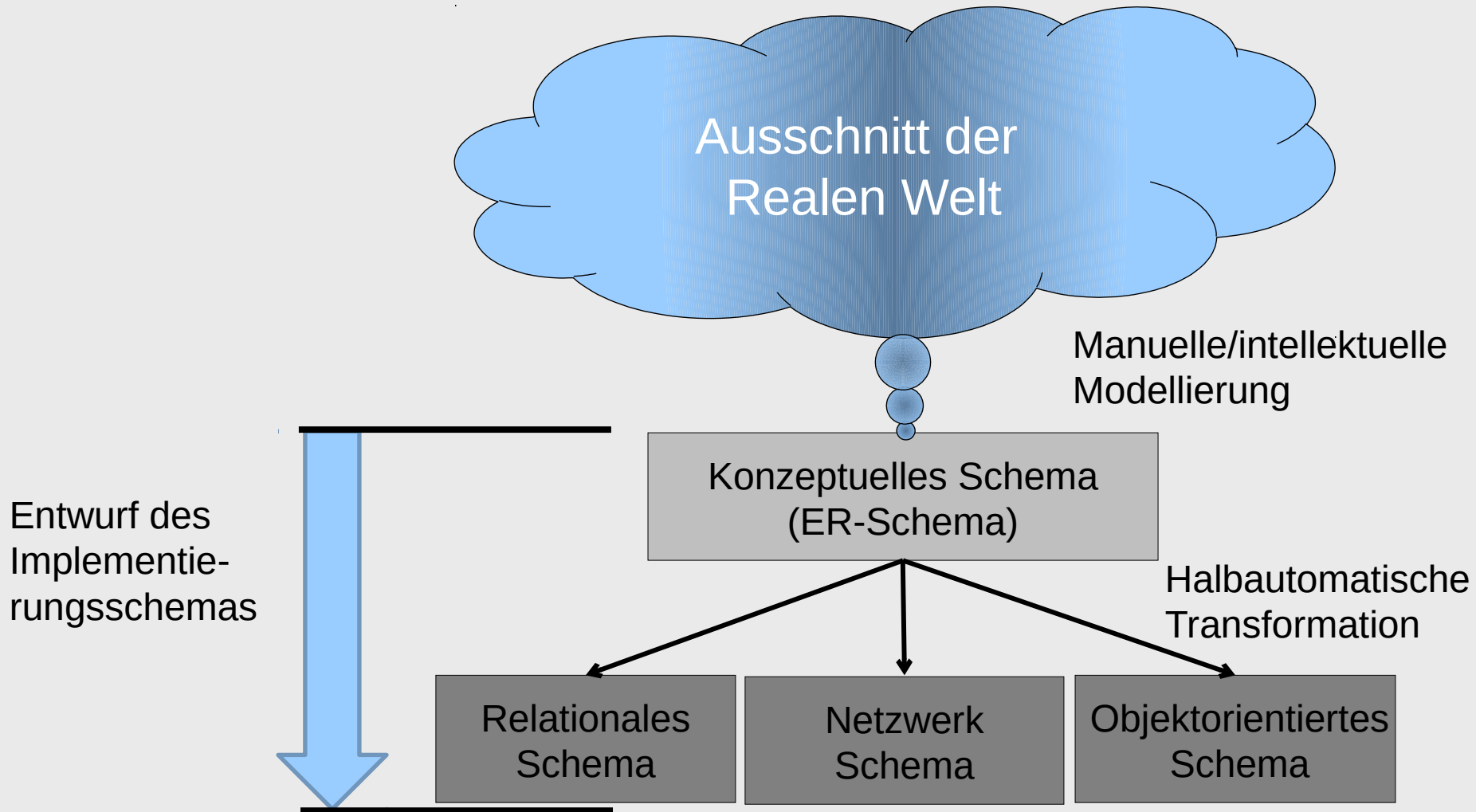
Konzeptuelle  
Modellierung



## Entity Relationship-Modell

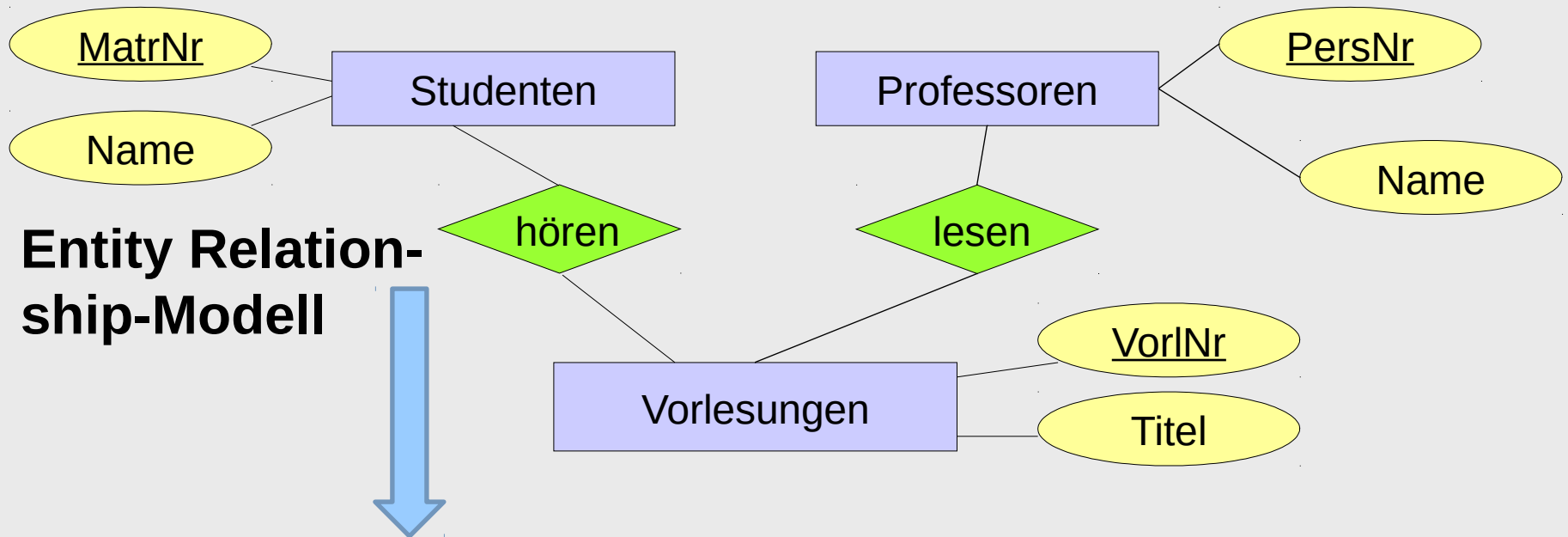


# Datenbankentwurf: Zweiter Schritt





# Beispiel Implementierungsschema



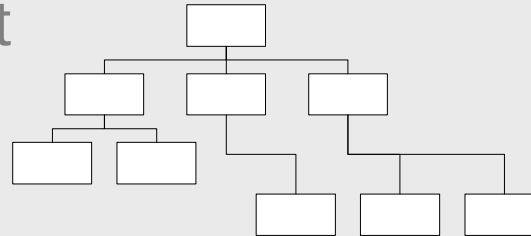
**Entity Relation-  
ship-Modell**

**Relationales Modell** (oder ein anderes Modell (z.B. das OO))

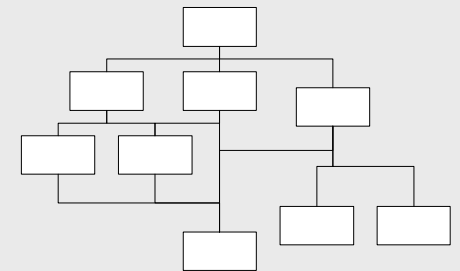
Studenten		hören		Vorlesungen	
MatrNr	Name	MatrNr	VorlNr	VorlNr	Titel
26120	Fichte	25403	5022	5001	Grundzüge
25403	Jonas	26120	5001	5022	Glaube und Wissen
...	...	...	...	...	...

# Schemakonzepte

- Hierarchisches Konzept



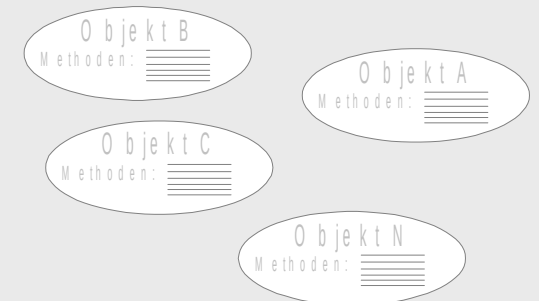
- Netzwerk-Konzept



- Relationales Konzept

	F · N R		F · N R	K · N R
	1		1	1
			1	2
			1	3
			1	4
	K · N R			
	1			
	2			
	3			
	4			

- Objektorientiertes Konzept



# Abbildung in einem relationalen Schema

Spalte mit Name und Werten				
Tabellenname				
Zeichnung	AUFTRAG	BEARBEITER	BENENNUNG	ID (Primary Key)
	Abt. B3	Weber	Pumpengehäuse	1
	Abt. A16	Weber	Ventildeckel	2
	Abt. B12	Springer	Layoutskizze	3
	Abt. B3	Meister	Gehäusedeckel	4
	Abt. B3	Springer	Flansch	5
	Abt. B12	König	Stromplan	6
	Abt. A16	Meister	Nockenwelle	7
	Abt. A5	Meister	Getriebekasten	8
DATENSATZ	Abt. A5	Weber	Hauptlager	9
	.....	.....	.....	

- **Schema:** Tabelle mit dem Namen „Zeichnung“ und den Spalten „Auftrag“, „Bearbeiter“, „Benennung“ sowie „ID“
- **Instanz:** Die gespeicherten Datensätze

# Nutzung von Daten

Wenn „Datenbank“ verfügbar:

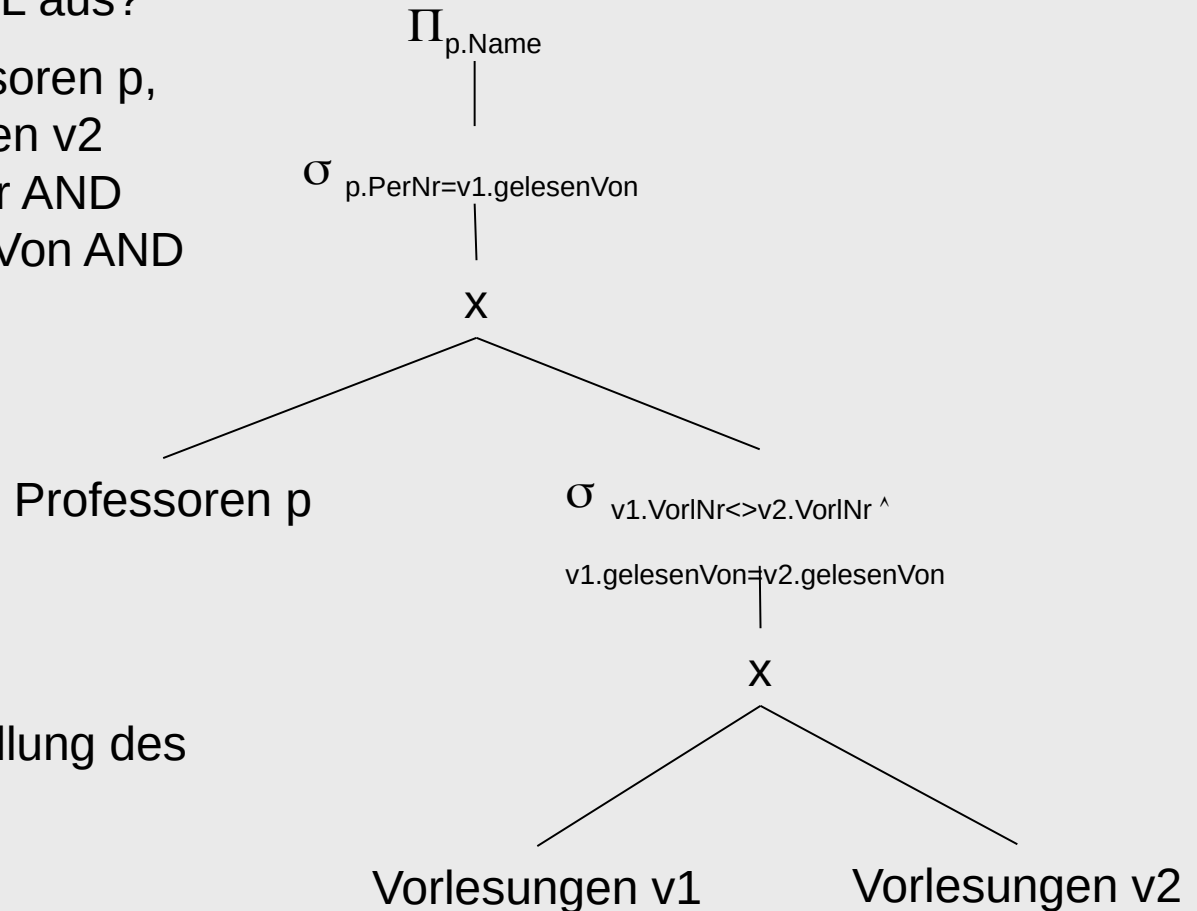
- Einfügen von Daten,
- Ändern von Daten,
- Abfrage der Daten,
- Verknüpfung von unterschiedlichen Daten,
- Aggregation von Daten
- Löschen von Daten
- ....

Wird mittels Structured Query Language (SQL) bewältigt.

# Relationale Algebra und Operatorbäume

Wie sieht die Abfrage in SQL aus?

```
select p.Name from Professoren p,
Vorlesungen v1, Vorlesungen v2
where v1.VorlNr<>v2.VorlNr AND
v1.gelesenVon=v2.gelesenVon AND
p.PersNr=v1.gelesenVon
```



Wie sieht die Inline-Darstellung des Operatorbaums aus?

$$\Pi_{p.Name}(\sigma_{p.PersNr=v1.gelesenVon}(\text{Professoren } x (\sigma_{v1.VorlNr \neq v2.VorlNr \wedge v1.gelesenVon=v2.gelesenVon}(\text{Vorlesungen } v_1 \times \text{Vorlesungen } v_2))))$$

## Weitere Themen

- Sichten
- Zugriffsrechte
- Qualität von Datenbankschemata
- Transaktionen
- Anbindung mit Java (JDBC)
- Mapping von Objekten zu Tabellen
- ...

# Heutige Datenbankgrößen

- eBay Data Warehouses 6,5PB =  $6,5 \cdot 10^{15}$  Bytes
- WalMart Data Warehouse 2,5 PB
- Facebook 400TB

Zum Vergleich:

- US Library of Congress 10-20TB  
(nicht digitalisiert)

Zahlen ca. 8 Jahre alt!!!!

Heute(2019): Facebook 4 Petabyte Daten fallen täglich an!

# Datenbankentwurf: Physische Ebene

Optimierung der Datenbank auf Speicherebene

- Indexstrukturen z.B. Suchbäume, Hashverfahren
- Replikation
- Physische Ebene ist unabhängig von der Implementationsebene

**Früher in Datenbanken II – teilweise Stoff des  
Schwerpunkts Soziale Webtechnologien**



# Datenbankentwicklungen heute

## **Unterstützung für spezielle Anwendungen:**

- Hochskalierbare, parallele Datenbanksysteme: Umgang mit Datenmengen im PB-Bereich
- Datenstromverarbeitung: Online-Verarbeitung von Live-Daten (Börseninfos, Sensordaten, RFID-Daten, . . . )
- XML-/JSON-Datenbanken: Verwaltung semistrukturierter Daten
- Multimediadatenbanken: Verwaltung multimedialer Objekte (Bilder, Audio, Video)
- Föderierte Datenbanken, Multidatenbanken, Mediatoren: Integration von Daten aus heterogenen Quellen (Datenbanken, Dateien, Web-Quellen)
- Mobile Datenbanken: Datenverwaltung auf Kleinstgeräten (PDA, Handy, . . . )

Vielen Dank für

---

Ihre Aufmerksamkeit

Prof. Dr.-Ing. Hendrik Gärtner