



Datenbanken und Informationssysteme

VL 01, Einführung in Datenbanken und Informationssysteme

SoSe 2022

Univ.-Prof. Dr.-Ing. habil. Norbert Gronau
Lehrstuhlinhaber | Chairholder

Mail August-Bebel-Str. 89 | 14482 Potsdam | Germany
Visitors Digitalvilla am Hedy-Lamarr-Platz, 14482 Potsdam
Tel +49 331 977 3322

E-Mail ngronau@lswi.de
Web lswi.de

Organisation der Lehrveranstaltung

Vorlesung

- Wöchentliche Vorlesung
- Vermittelt theoretische und methodische Grundlagen von Datenbanken und Informationssystemen
- Montag 10:15 - 11:45 Uhr
- In Präsenz: 3.06.H02
- Materialien zur Lehrveranstaltung werden auf der Homepage des Lehrstuhls im Lehrportal bereitgestellt

Übung

- Beginn am 25. April
- Eine Übungsgruppe:
- Montag 12:15 - 13:45 Uhr
- Findet nicht wöchentlich statt
- Via Zoom und Moodle (wird über Lehrportal bereitgestellt)
- Vermittlung praktischer Kenntnisse
- Erarbeitung und Besprechung der Übungsblätter

Organisatorische Hinweise - Übungsgruppenleiter



Übungsleiter

- Dr. rer. pol.
Edzard Weber
- Sprechstunde: n.V.
- Edzard.weber@
wi.uni-potsdam.de



Organisation

- M.Sc.
Marcel Rojahn
- Sprechstunde: n.V.
- Marcel.rojahn@
wi.uni-potsdam.de

Bitte denken Sie daran, sich auch über das Lehreportal des Bereichs Wirtschaftsinformatik Prozesse und Systeme anzumelden.

Prüfungsrelevante Studienleistungen

Klausur

- 100% der Gesamtnote
- Termin für die Klausur:
11. August / 8:15 -9:45 / 3.06 H02

Hausaufgabe

- Bearbeitung der Übungsblätter
- Weitere Informationen folgen in der Übung

Für den erfolgreichen Abschluss der Lehrveranstaltung ist das Bestehen der Klausur erforderlich! Zusätzliche Informationen zur Hausaufgabe erfolgen in der ersten Übung.

Inhalt

1. Einführung in Datenbanken und Informationssysteme
2. Physische Datenorganisation
3. Relationales Datenmodell
4. Hierarchisches Datenmodell
5. Objektorientiertes Datenmodell
6. Sicherheit
7. Business Intelligence: Data Warehousing
8. Data Mining
9. Big Data Analytics
10. Big Data Engineering Teil 1
11. Big Data Engineering Teil 2

Lernziele der Vorlesung DBIS

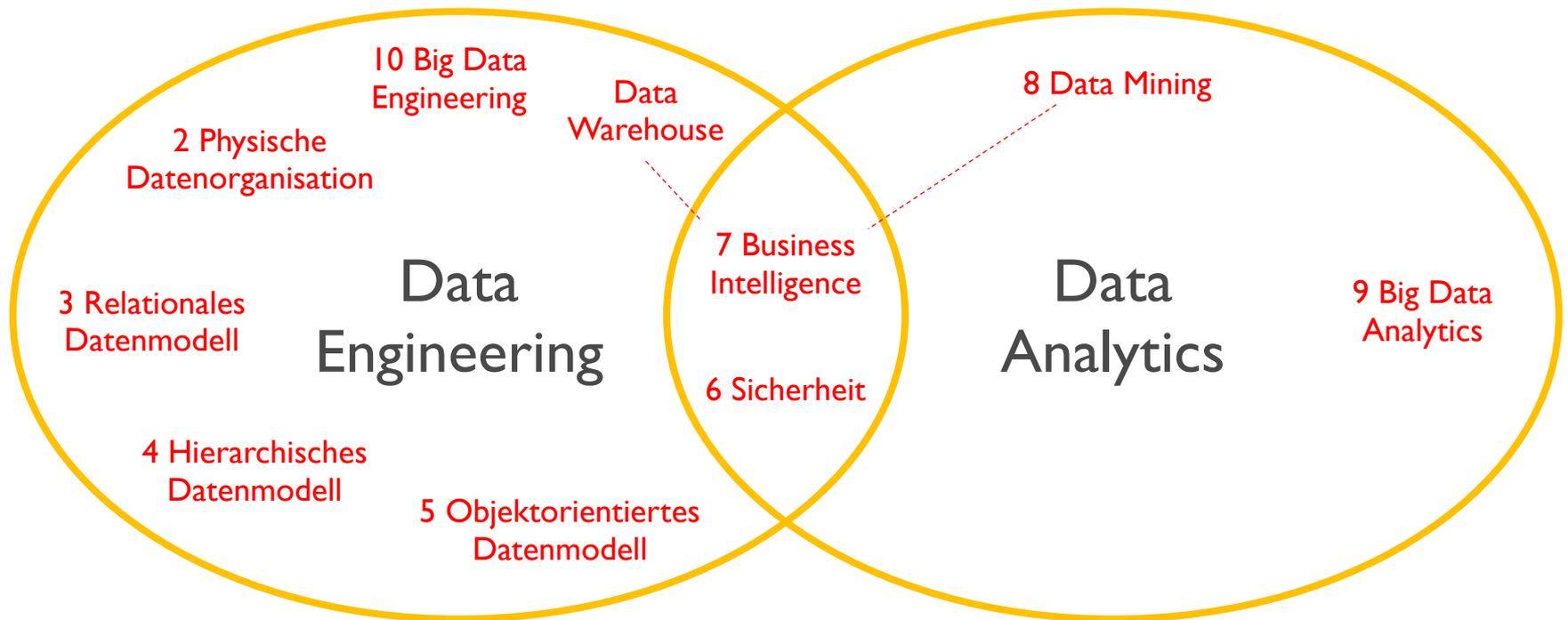
■ Wissen

- > Physische Datenorganisation, Relationales Datenmodell (SQL), XML, UML, Sicherheit, Big Data, NoSQL, In-Memory-DB
- > Probleme und Determinanten des praktischen Einsatzes

■ Können

- > Professionalisierung ERD/SQL
- > Anwendung aller Modellierungstechniken
- > Aber vor allem: *Routine* im Umgang mit Daten!

Teilbereiche der DBIS



Kapitel I: Einführung

■ ■ Begriffe

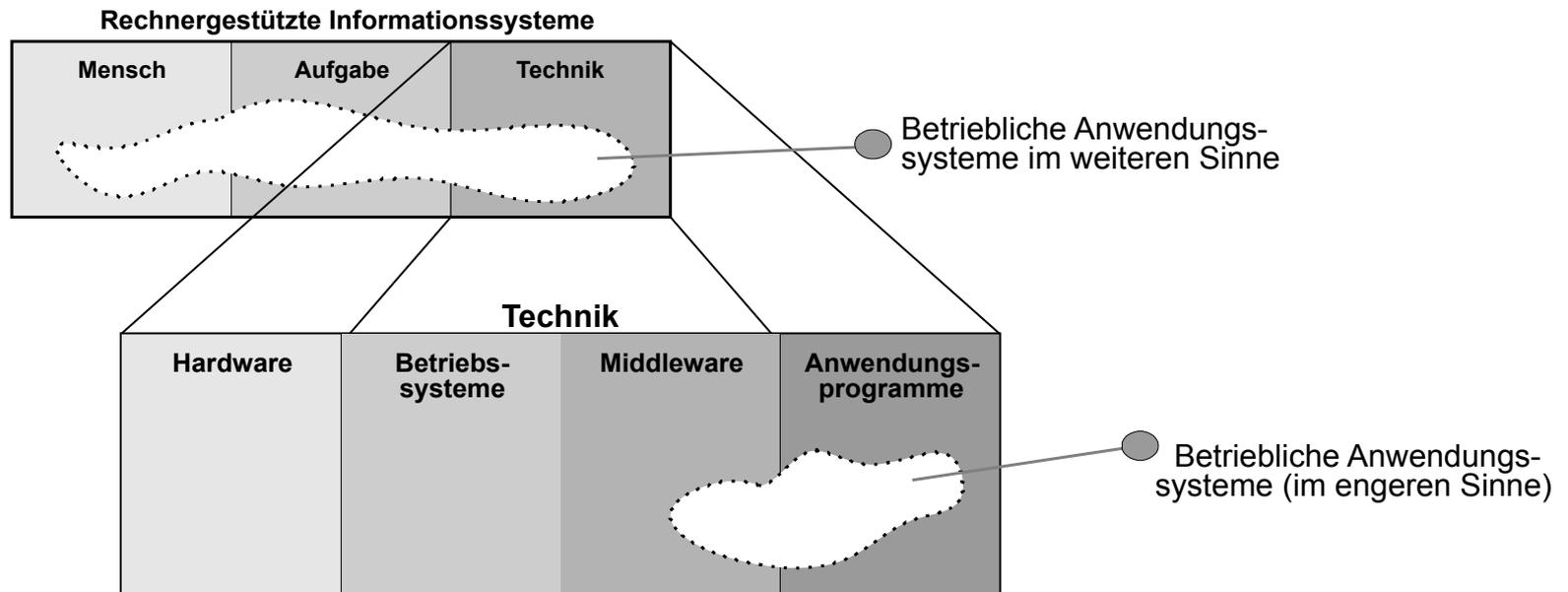
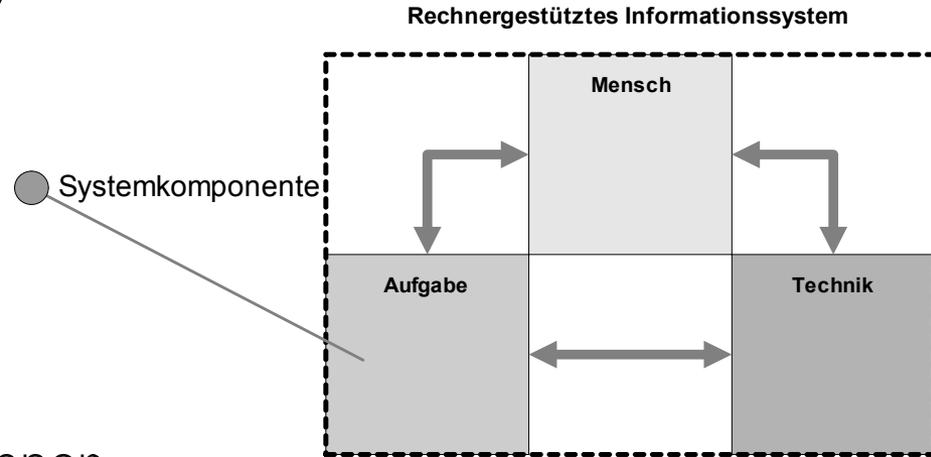
- Wissen
 - > Gesamtheit der Kenntnisse auf einem Gebiet
- Information
 - > handlungsrelevantes Wissen
- Daten
 - > Abbildung von Information in Form von Zeichen oder kontinuierlichen Funktionen zum Zwecke der Verarbeitung
 - > heute typischerweise in digitaler Form vorliegend

Informationssysteme

- Informationssystem (IS)
 - > umfasst Mittel, Maßnahme und Einrichtung für die Produktion, Verbreitung und Nutzung von Informationen im Unternehmen, um die Mitarbeiter mit Informationen zu versorgen, die diese zur Aufgabenerfüllung benötigen
- Heute typischerweise rechnergestützt
 - > Synonym: Informations- und Kommunikationssystem
- In der WI versteht man ein IS generell als *Mensch-Aufgabe-Technik-System*

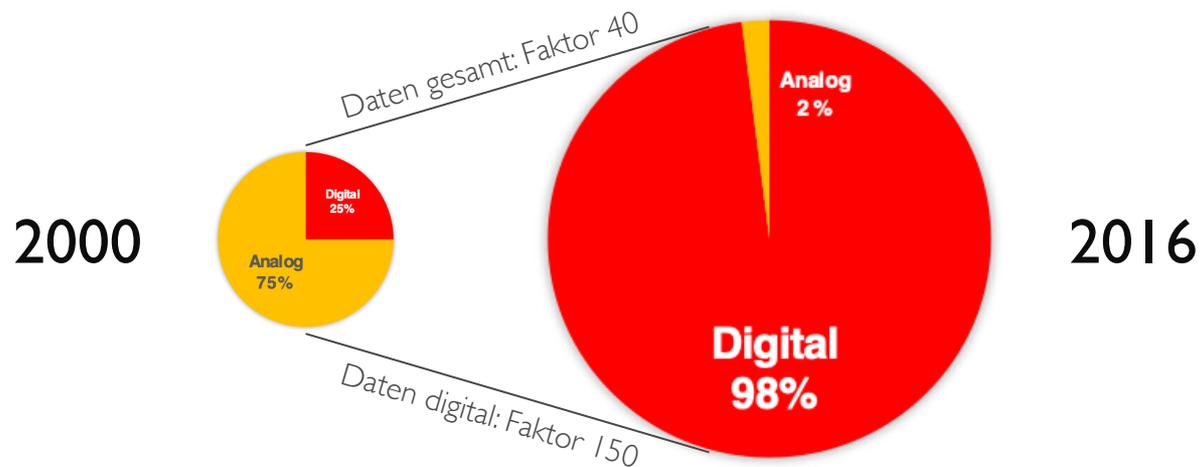
Mensch-Aufgabe-Technik-System

- Systemkomponenten
 - > Mensch, Aufgabe, Technik
- Anwendungssystem
 - > Teil der Technik-Komponente
 - > Häufig auch Synonym für rechnergestützte IS zur (Teil-)Automatisierung betrieblicher Aufgaben



"Datafication"

- Zunehmende Erfassung und Digitalisierung von Daten
- Quantifizierung zahlreicher Aspekte des Alltags, der Wirtschaft und des Lebens durch Big Data



Beispiele für Datafication

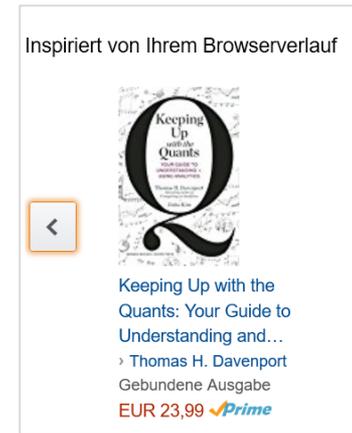
- Konsumverhalten
 - > z.B. Apps, Sensoren, Weblogs, mobiles Bezahlen
- Privatleben
 - > z.B. Apple Siri, Amazon Echo, Facebook, WhatsApp
- Industrie
 - > z.B. RFID, Cyberphysische Systeme
- Städte
 - > z.B. Kameras, IoT
- Gesundheitswesen
 - > z.B. Smartwatches, Health-Tracker



Beispiel: Empfehlungssystem von Amazon



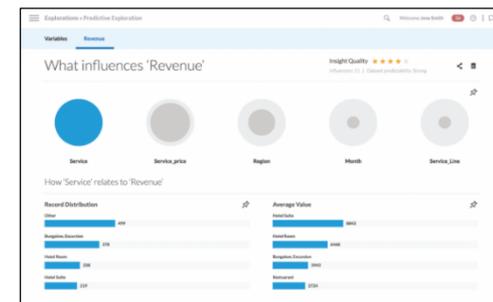
- Vielfältige Fragestellungen zum Kundenverhalten
 - > Was schauen sich Kunden an?
 - > Wie sind sie durch die Webseite navigiert?
 - > Wie sehr lassen sie sich durch Promotionen, Bewertungen und Seitenlayouts beeinflussen?
 - > Welche Ähnlichkeiten gibt es zwischen den Individuen und Gruppen?
- Echtzeitauswertung der Daten
 - > auch: sehr große, unstrukturierte Datenmengen (Clickstreams)
- Anwendung von Big-Data-Algorithmen
 - > Prognose und Einflussnahme ("passgenaue Angebote")
 - > Warenverfügbarkeit und Lagerhaltung



Beispiel: SAP Digital Boardroom



- Echtzeit-Berichte über finanzielle, operative, marketing- und vertriebsrelevante Kennzahlen
- Vergleich der möglichen Auswirkungen unterschiedlicher Geschäftsstrategien durch *Was-wäre-wenn-Analysen*
- Grundlage: In-Memory-Datenbank *S/4 HANA*



Datenkategorisierung (I)

Unterscheidungsmerkmal	Ausprägung	Erklärung bzw. Bsp.
<i>Funktion</i>	Primärfunktionsdaten	Daten von/für Beschaffung, Produktion, Vertrieb/Absatz, Kunden, Finanzen
	Sekundärfunktionsdaten	Daten von/für Organisation, Planung, Personal, Verwaltung, Logistik, FuE
<i>Anwendung</i>	Stammdaten	wichtige Grunddaten eines Betriebes, die längerfristig unverändert bleiben (z.B. Kontonummer)
	Bewegungsdaten	Daten zur Aktualisierung von Stamm-/Bestandsdaten und Beschreibung von Veränderungen
	Bestandsdaten	zustandsorientierte, regelmäßig aktualisierte Daten eines Betriebes (z.B. Kontensaldo)
<i>Verarbeitung</i>	Eingabedaten	werden Programm von außen zur Verfügung gestellt
	Ausgabedaten	stellen Ergebnis eines Programms dar
<i>Zeitpunkt</i>	Realtime-Daten	Originaldaten, die in in Echtzeit zur Verfügung stehen (z.B. Maschinendaten)
	Neartime-Daten	Kopiedaten mit etwas geringerer Aktualität als die Originaldaten
	Historische Daten	Datenbestand zu bestimmten Zeitpunkten, der getrennt gespeichert und aufgerufen wird (z.B. Jahresanfangsbestand auf Monitor angezeigt)

Datenkategorisierung (2)

Unterscheidungsmerkmal	Ausprägung	Erklärung bzw. Bsp.
<i>Neuartigkeit</i>	Originäre Daten	unverarbeitete Daten (originär), so wie sie anfallen
	Abgeleitete Daten	aufbereitete, weiterverarbeitete Daten von/für Organisation, Planung, Personal, Verwaltung, Logistik, FuE
<i>Dimensionalität</i>	Eindimensionale Daten	univariate Daten: pro Objekt wird ein Merkmal durch Messung / Befragung / Beobachtung erhoben
	Mehrdimensionale Daten	multivariate Daten, Mehrattributanfragen, geographische Informationssysteme
	Hochdimensionale Daten	(Farb-)Histogramme über Multimedia-Daten, Zeitreihen, Messreihen
<i>Strukturierung</i>	Strukturierte Daten	Daten mit gleichartiger Struktur in einem Datenbankmodell (z.B. Tabellen)
	Semistrukturierte Daten	Daten mit gewisser aber nicht sofort erkennbarer Struktur die nicht in relationale oder objektorientierte Datenbank-schemata passen (z.B. formatierte Texte, XML, HTML)
	Unstrukturierte Daten	digitalisierte Informationen ohne Struktur, die nicht über eine einzelne Schnittstelle verarbeitet werden können (z.B. Tonaufnahmen, Texte, Grafiken)

Stufen der Datenanalyse nach Davenport

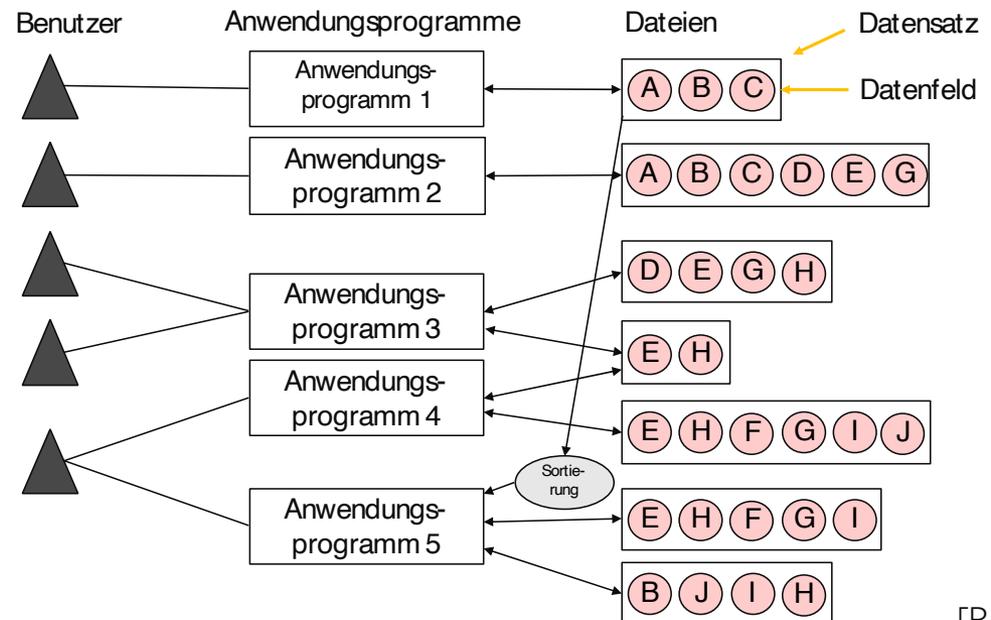
	Analytics 1.0	Analytics 2.0	Analytics 3.0
Types of companies	Large enterprises	Online and start-ups	All—“data economy”
Analytics objective	Internal decisions	New products	Decisions and products
Data type	Small, structured	Large, unstructured	All types combined
Creation approach	Long-cycle batch	Short-cycle agile	Short-cycle agile
Primary technology	Software packages	Open source	Broad portfolio
Primary analytics type	Descriptive	Descriptive, predictive	Prescriptive
Business relationship	Back office	“On the bridge”	Collaborative

Alte Welt der Programmierung

- Programm im Fokus, Daten inkludiert oder in separaten Dateien
 - > enge Verflechtung aus Programmentwurf und Datenorganisation auf den Speichermedien
 - > für jede Anwendung eigene Dateien mit erforderlichen Datensätzen und spezifischen Zugriffsfunktionen

■ Probleme

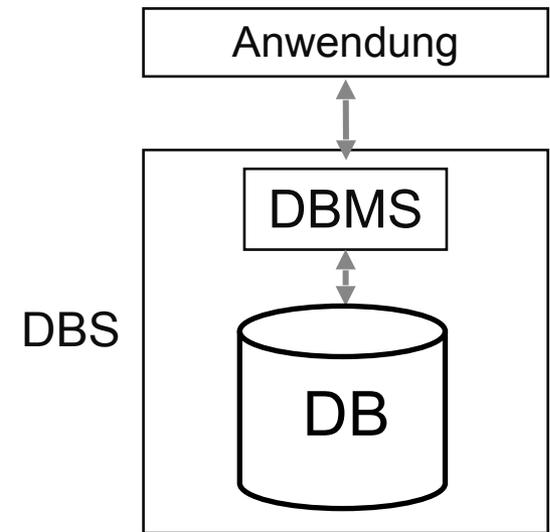
- > Redundanz
- > Inkonsistenz
- > hoher Aufwand



[RePi9 I]

Neue Welt: Datenbanken

- Datenbanksystem (DBS)
 - > Werkzeug zum Umgang mit dauerhaft gespeicherten, großen, heterogenen Datenmengen
- Datenbank (DB)
 - > logisch integrierter Datenbestand, der unter Verwendung eines DBMS verwaltet wird und mit Hilfe eines *Datenmodells* beschrieben ist
- Datenbankmanagementsystem (DBMS)
 - > Werkzeug zur Einrichtung, Verwaltung und Nutzung einer Datenbank durch Anwendungen oder Einzelbenutzer mittels standardisierter Schnittstellen



Entwicklung

- 1960er: erste kommerzielle DBS
 - > hervorgegangen aus großen Dateisystemen mit ineffizientem Zugriff und geringer Sicherheit
 - > Anwendung in Transaktionssystemen
 - » z.B. im Bankwesen, für Flugreservierungssysteme oder in der Unternehmensbuchhaltung (Löhne, Bestände, Verkäufe)
 - > Datenmodelle zur Strukturbeschreibung ohne höhere Abfragesprachen
 - » Netzwerkdatenmodell oder hierarchisches Datenmodell
- 1970: Datensicht organisiert in Tabellen (Relationen)
 - > Structured Query Language (SQL) als höhere Abfragesprache
 - > seit Anfang der 1990er Jahre weitgehend etabliert

Trends

■ Trend 1: Immer kleiner werdende Systeme

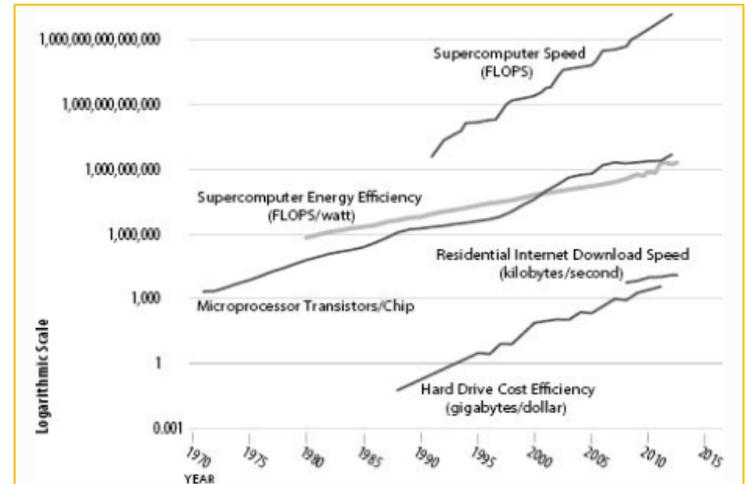
- > relationale Datenbanksysteme etablierten sich und wurden verfügbar für kleinere Rechnersysteme
- > große Sammlungen von kleinen Dokumenten in Datenbanken durch das Internet, z.B. XML (eXtensible Markup Language)

■ Trend 2: Immer größer werdende Systeme

- > stetige Zunahme der Datenmengen und Dateigrößen
 - » Bilddaten (z.B. Flickr, Instagram)
 - » Online-Konsum von Videos (z.B. Youtube, Netflix)
 - » Netzwerke zur verteilten Datenspeicherung (z.B. Filesharing in Peer-to-Peer-Systemen)
- > Relationale DBS für unstrukturierte Daten eher ungeeignet
 - » Neuentwicklungen im Bereich paralleler Datenverarbeitung (z.B. Hadoop), NoSQL/NOSQL-Datenbanken

Moore's Law

- Verdopplung der Leistungsparameter alle 18 Monate
 - > Prozessorgeschwindigkeit
 - > Hauptspeicherkosten pro Bit
 - > Mikrochipdichte
 - > Diskkosten pro Bit
 - > Kapazität der größten Disks



- Einige Parameter folgen jedoch nicht dieser Regel
 - > Zugriffsgeschwindigkeit im Arbeitsspeicher
 - > Rotationsgeschwindigkeit von Festplatten
- Folge: Zugriffszeit sinkt nicht proportional
 - > Bewegung von Daten innerhalb der Speicherhierarchie im Vergleich zur Prozessorgeschwindigkeit

Beispiele für Datenbankgrößen

How big is a Yottabyte?

TERABYTE	Will fit 200,000 photos or mp3 songs on a single 1 terabyte hard drive.	
PETABYTE	Will fit on 16 Backblaze storage pods racked in two datacenter cabinets.	
EXABYTE	Will fit in 2,000 cabinets and fill a 4 story datacenter that takes up a city block.	
ZETTABYTE	Will fill 1,000 datacenters or about 20% of Manhattan, New York.	
YOTTABYTE	Will fill the states of Delaware and Rhode Island with a million datacenters.	

Big Data at Big Companies

Google processes **3.5 billion** requests per day. This requires a massive amount of data storage. It is estimated that Google is one of the largest big data companies in the world given their many products and functions. In 2009, Google was preparing to have over **1 exabyte** of data. It is estimated that Google is now storing over 10 exabytes of data.

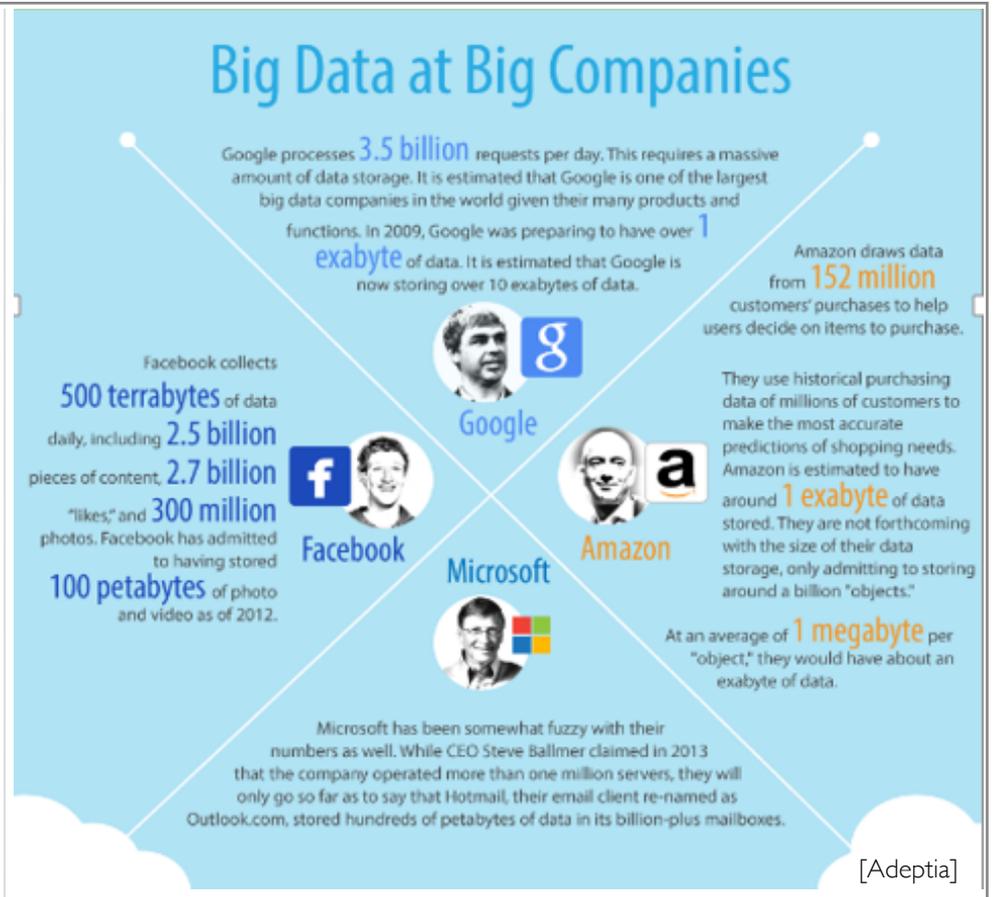
Amazon draws data from **152 million** customers' purchases to help users decide on items to purchase. They use historical purchasing data of millions of customers to make the most accurate predictions of shopping needs. Amazon is estimated to have around **1 exabyte** of data stored. They are not forthcoming with the size of their data storage, only admitting to storing around a billion "objects."

At an average of **1 megabyte** per "object," they would have about an exabyte of data.

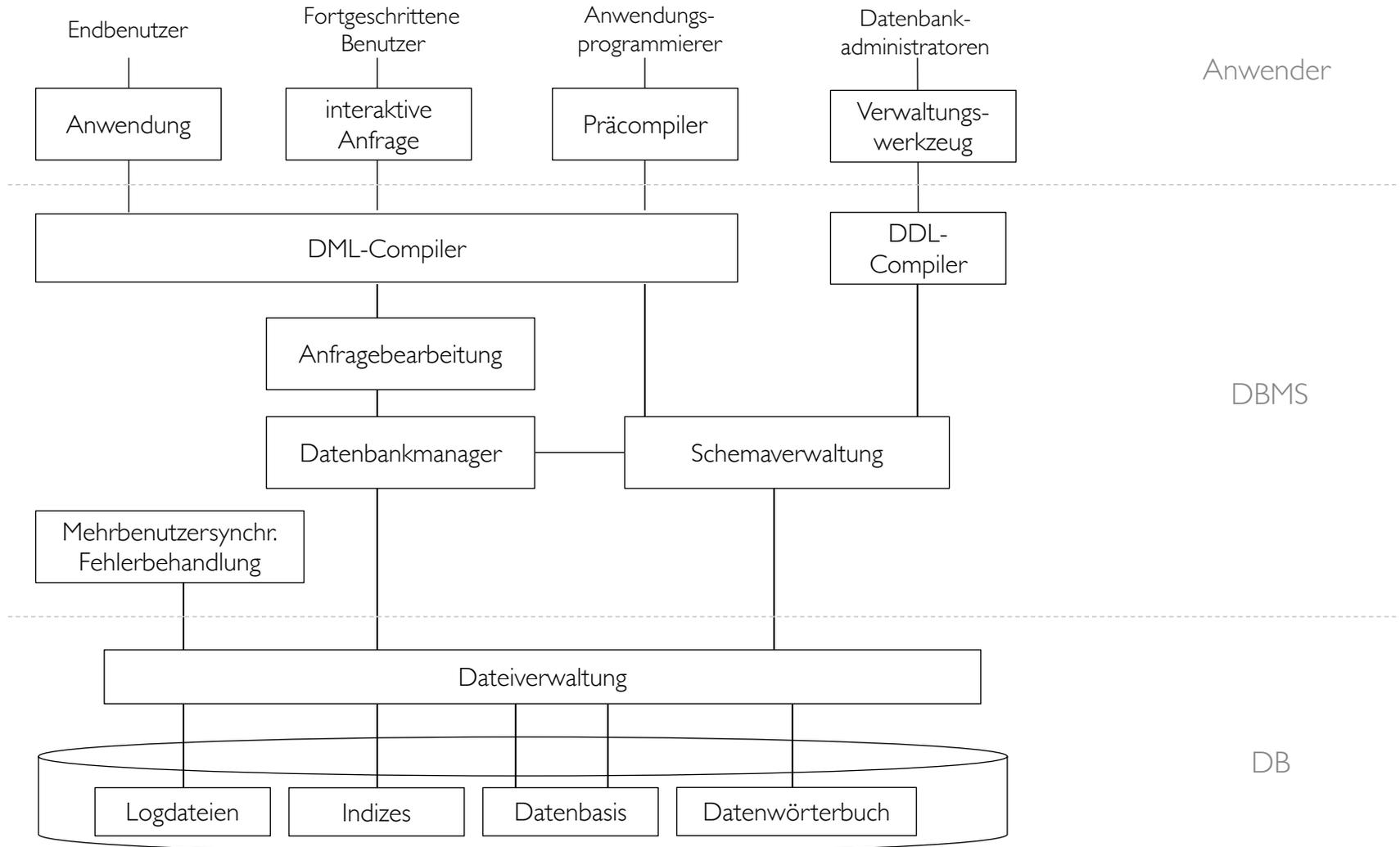
Facebook collects **500 terabytes** of data daily, including **2.5 billion** pieces of content, **2.7 billion** "likes," and **300 million** photos. Facebook has admitted to having stored **100 petabytes** of photo and video as of 2012.

Microsoft has been somewhat fuzzy with their numbers as well. While CEO Steve Ballmer claimed in 2013 that the company operated more than one million servers, they will only go so far as to say that Hotmail, their email client re-named as Outlook.com, stored hundreds of petabytes of data in its billion-plus mailboxes.

[Adeptia]



Architekturübersicht eines DBS



[KeEi15]

Anwendersicht

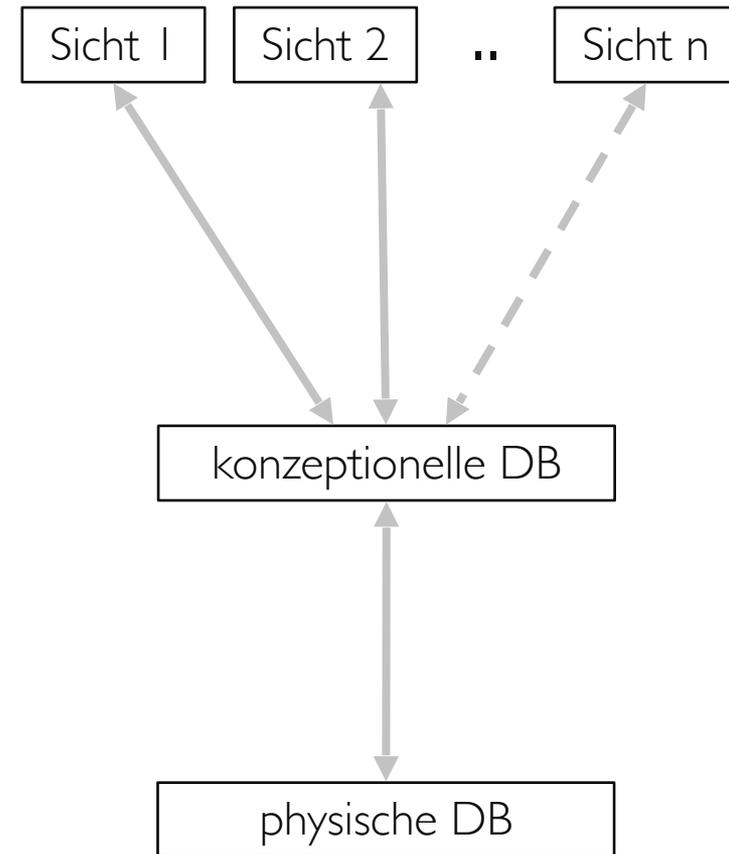
- Anwendungssysteme/-programme
 - > speziell abgestimmte Programme für Endbenutzer zur Durchführung repetitiver Aufgaben
- Interaktive Anfragen
 - > Tools und Programme zur Durchführung komplexerer Anfragen für fortgeschrittene Benutzer
- Präcompiler
 - > Einbettung von Elementen der Abfragesprache in eine Programmiersprache zur Durchführung besonders komplexer Datenverarbeitungsanforderungen für Programmierer
- Verwaltungswerkzeuge
 - > für Datenbank-Administratoren zur Schema-Manipulation und Benutzerkennungs-Verwaltung

Komponenten des DBS

- DML-Compiler
 - > Untersuchung der Anforderungen von Daten und Transformation in eine für die Anfragebearbeitung verständliche Form
- Anfragebearbeitung
 - > Prüfung auf effiziente Erfüllung der Anforderungen und Umwandlung in Unterprogrammaufrufe des Datenbankmanagers
- Datenbankmanager
 - > Kernstück des DBMS und die Schnittstelle zur Dateiverwaltung
- DDL-Compiler
 - > Analyse der Schemamanipulationen durch den DB-Administrator und Übersetzung in Metadaten
- Schemaverwaltung
 - > Verarbeitung der Metadaten und Speicherung im Datenwörterbuch
- Mehrbenutzersynchronisation/Fehlerbehandlung
 - > Anlegen von Archivkopien und Protokollen

Datenabstraktion: ANSI Drei-Ebenen-Modell

- Externes Schema (Sichten, Views)
 - > Beschreibung der zu verwaltenden Daten (und deren Beziehungen) aus der Sicht des (einzelnen) Anwenders
 - » Welche Daten werden für Aufgabe benötigt?
- Konzeptionelles Schema
 - > logische Gesamtbeschreibung der zu verwaltenden Daten sowie deren Beziehungen untereinander
 - » Welche Daten sind gespeichert?
 - » Wie stehen die Daten jeweils in Beziehung?
- Internes Schema
 - > formale Beschreibung, wie (und mit welcher Technik) die zu verwaltenden Daten abzuspeichern sind und wie auf diese zugegriffen werden kann
 - » Wie sind die Daten tatsächlich gespeichert?



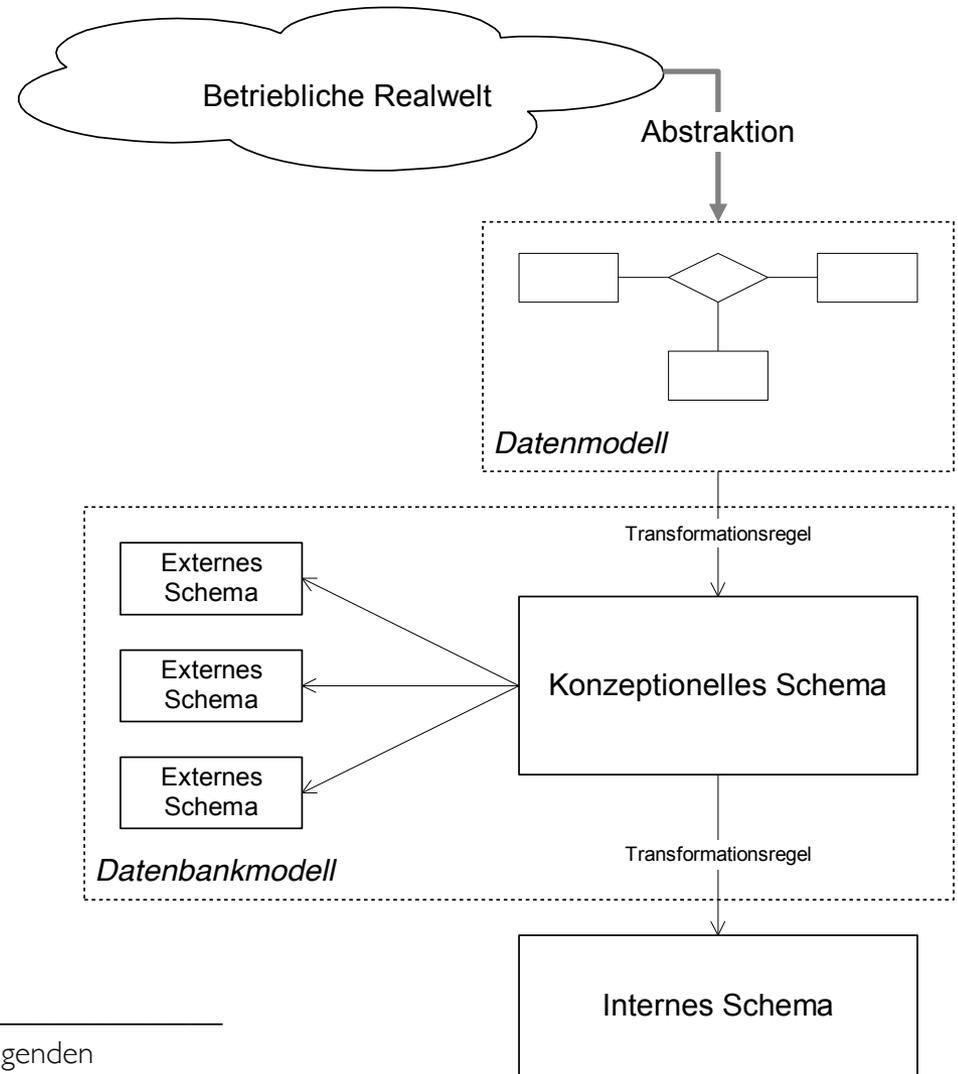
Daten- und Datenbankmodelle

■ Datenmodell

- > Beschreibung fachlicher Anforderungen des in der Datenbank abzubildenden Realitätsausschnitts

■ Datenbankmodell¹

- > Konkrete, datenbank-systembezogene Umsetzung des Datenmodells



¹ Begriff wird nur noch selten verwendet und daher im Folgenden unter dem Begriff "Datenmodell" subsumiert

Forderungen an Datenbanken

- Konsistenz

- > Daten sind widerspruchsfrei

- Korrektheit

- > Daten entsprechen der Realität

- » inkonsistente Daten können nicht korrekt sein

- Integrität

- > korrekter, widerspruchsfreier, vollständiger Zustand der Daten

- Redundanz

- > Daten werden an verschiedenen Stellen gespeichert

- » erwünscht: Sicherungen, Erhöhung der Ausfallsicherheit, Performance

- » unerwünscht: Doppelte Pflege der Daten → Konsistenzproblem

 Vermeidung von Redundanz, soweit möglich
(Ggf. Kompromisse hinsichtlich Sicherheit oder Performance)

■ Fragestellungen zur Datenerfassung

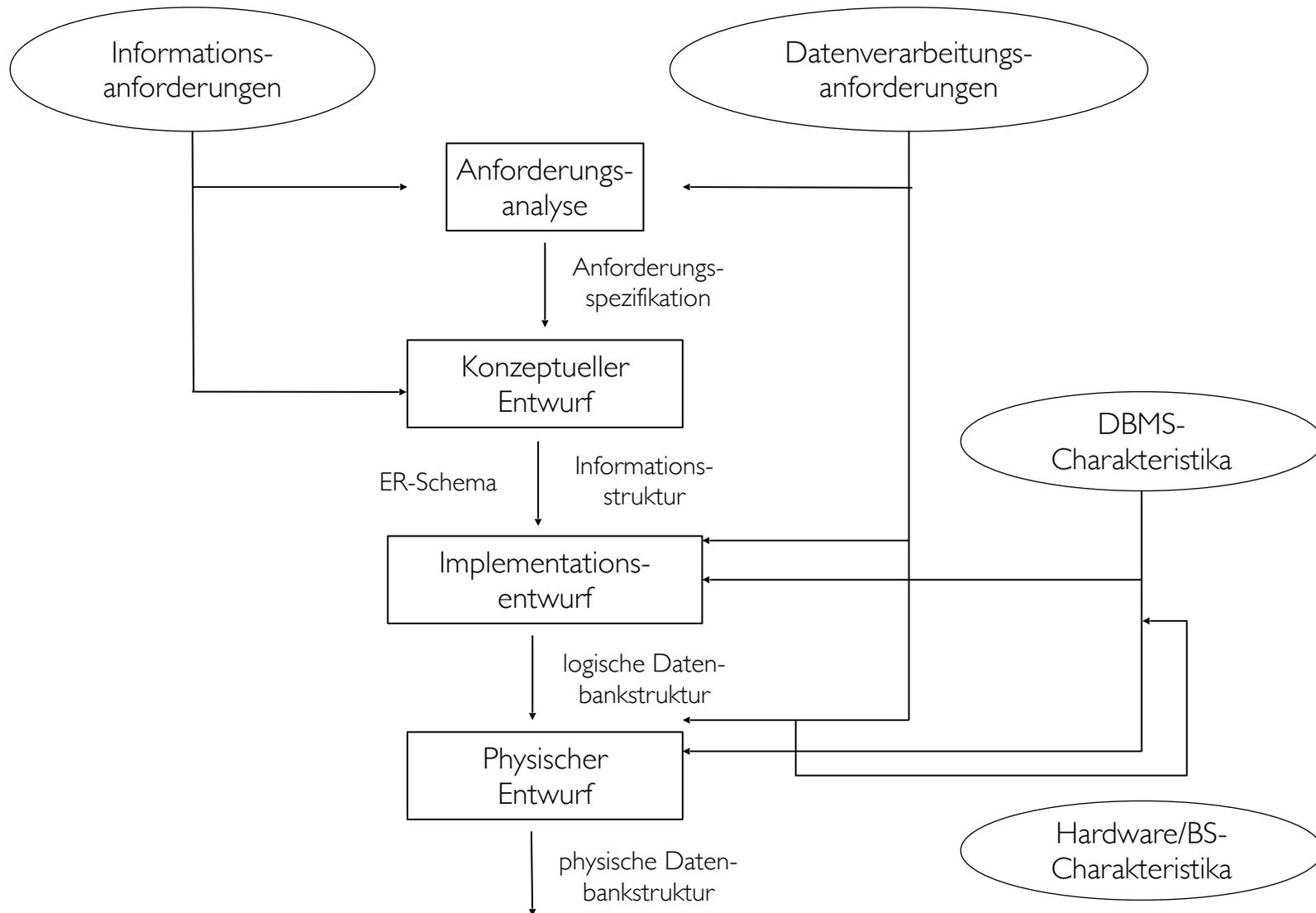
■ Datenanfall

- > Ort (Quelle, Speicherort)
- > Zeit (Zeitpunkt, Dringlichkeit, Intervalle)
- > Vielfalt (Menge, Größe, Komplexität)
- > Form (Format, Kompatibilität)
- > Qualitätsanspruch (Vollständigkeit, Korrektheit, Konsistenz)
- > Beeinflussungsmöglichkeiten

■ Datenerfassung

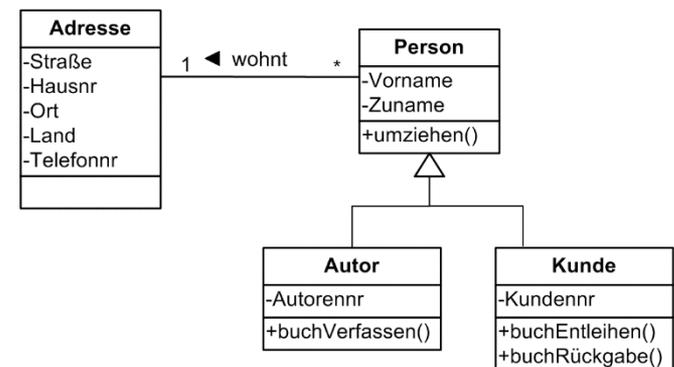
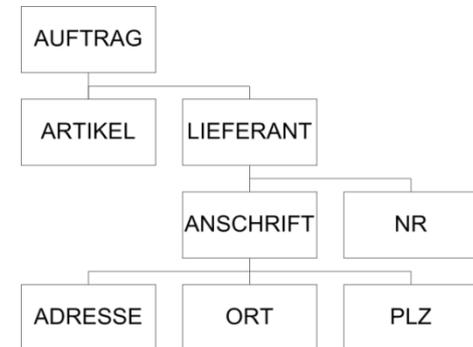
- > Verbindungsgrad von Erfassung und Verarbeitung skalierbar?
- > Codierungsfreiheit
 - » z.B.: Kfz-Kennzeichen statt Ortsnamen

Phasen des Datenbankentwurfs



Verschiedene Datenmodelle

- **Relationales Datenmodell**
 - > Datenobjekte und deren Beziehungen werden in Tabellen (=Relationen) gespeichert
 - > am häufigsten verwendetes Datenmodell
 - > Beispiele: Oracle, DB2, MySQL
- **Hierarchisches Datenmodell**
 - > Daten werden als *Baumstrukturen* dargestellt
 - > Beispiele: XML
- **Objektorientiertes Datenmodell**
 - > Objekte werden mit *Eigenschaften* und *Methoden* direkt in der Datenbank gespeichert
 - > Beispiele: Db4o, Versant, ZODB

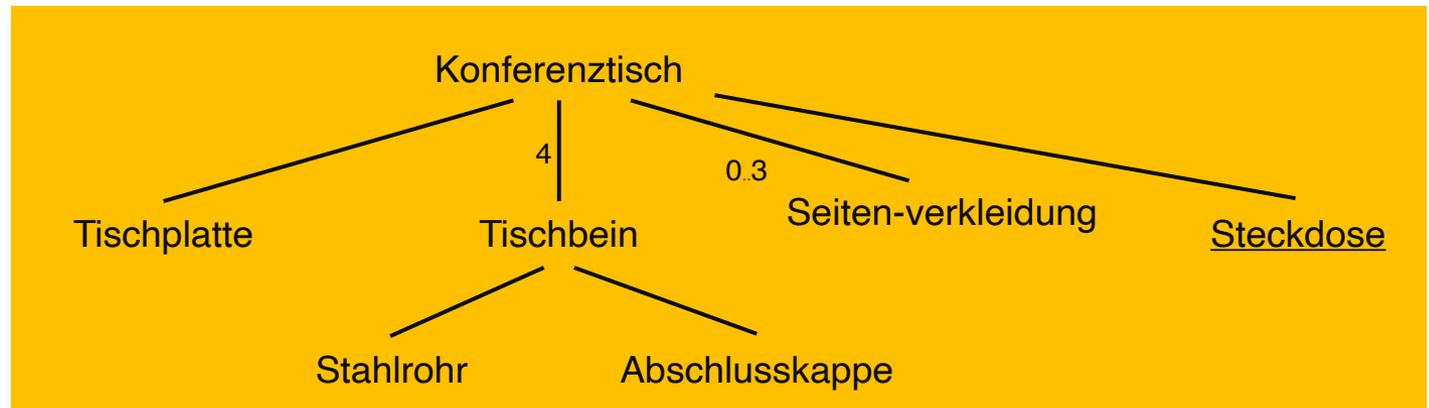


■ ■ Neue Entwicklungen bei Datenbanken

- NoSQL (i.S.v. "kein SQL")
 - > Datenbanken, die einen nicht-relationalen Ansatz verfolgen
- NOSQL (i.S.v. "Not only SQL")
 - > Datenbankumgebungen, die traditionelles SQL (oder SQL-ähnliche Abfragesprachen) mit alternativen Zugriffs- und Abfragemöglichkeiten kombinieren
 - > NewSQL: SQL mit hoher Skalierbarkeit
- No-SQL Frameworks
 - > Key-Value Stores
 - > Tabular-oriented Databases
 - > Document Stores
 - > Graph-oriented Databases
 - > Object Databases

P-Datenbaum

- Einfache Darstellungsmöglichkeit für ADT¹
 - > Bipartiter Graph
 - » Knoten (*Element* als alphanumerische Zeichenkette)
 - unterstrichen := Element ist getrennt dargestellter eigener ADT
 - » Kanten (bis in die unmittelbare Nähe des Knotens gezeichnet)
 - Beschriftung kann eine Anzahl von 2..n Nachfolgern vorgeben (sonst 1)
 - > Start: Stets ein Wurzelknoten mit dem Namen des ADT
 - > Ende: Stets Blattknoten
- Beispiel:

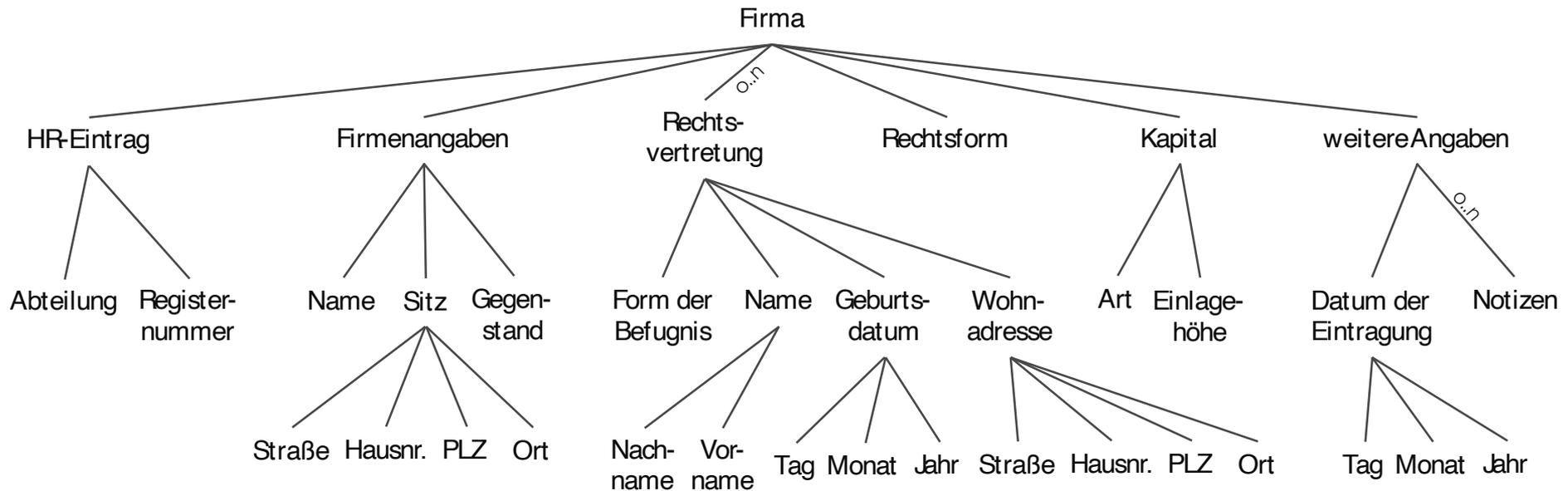


¹ oder eine Stückliste; dies gilt für jede Nennung des Begriffs ADT auf dieser Folie

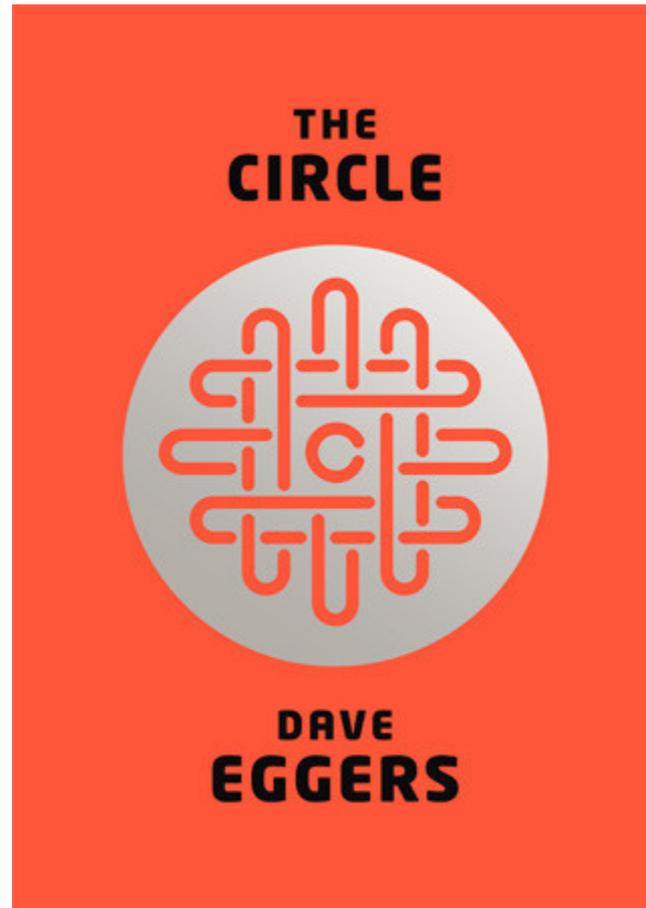
² stets links der Kante, etwa auf halber Höhe

Beispiel: Einfacher P-Datenbaum

Firma im Handelsregister



Exkurs I: Was man als WI-ler gelesen haben muss



Exkurs 2: Digitalisierung des Lernens

Turbo oder Produktivitätsvernichter?

> Lernverhalten

	nicht digital	digital
aktiv	V: Wichtiges mitschreiben Ü: Aufgaben selbst lösen E: Anwendungsbeispiele suchen	V: Wichtiges mitschreiben Ü: Aufgaben selbst lösen E: Anwendungsbeispiele suchen
passiv	V: Nur Zuhören/Durchlesen Ü: Lösungen nachvollziehen E: <i>-fehlt-</i>	V: Nur Zuhören/Durchlesen Ü: Lösungen nachvollziehen E: <i>-fehlt-</i>

> Wie gehen Führungskräfte mit der Digitalisierung um?

1. Entscheiden Sie sich bewusst, nicht aus dem Bauch heraus
2. Beobachten Sie Ihr Lernverhalten und Ihren Erfolg (schriftlich!)
3. Finden Sie die für Ihren Erfolg beste Kombination aus digitalem und nicht-digitalem Lernen heraus und halten diese konsequent durch

Literaturverzeichnis

Literaturverzeichnis

- [Bapp14] Bappalige, O. (2014). Introduction to Apache Hadoop.
URL: <https://opensource.com/life/14/8/intro-apache-hadoop-big-data>
- [BEPW16] Backhaus, K., Erichson, B., Plinke, W., Weiber, R. (2015). Multivariate Analysemethoden: Eine anwendungsorientierte Einführung. Springer-Verlag.
- [Dave14] Davenport, T. (2014). Big data at work: dispelling the myths, uncovering the opportunities. Harvard Business Review Press.
- [EFH+11] Edlich, S., Friedland, A., Hampe, J., Brauer, B., Brückner, M. (2011). NoSQL: Einstieg in die Welt nichtrelationaler Web 2.0 Datenbanken. Hanser.
- [FaEg05] Fassott, G., Eggert, A. (2005). Zur Verwendung formative und reflektiver Indikatoren in Strukturgleichungsmodellen, in: Handbuch PLS- Pfadmodellierung. Schäffer Poeschel.
- [FaPS96] Fayyad, U. M., Piatetsky-Shapiro, G., Smyth, P. (1996, August). Knowledge Discovery and Data Mining: Towards a Unifying Framework. In KDD (Vol. 96, pp. 82-88).
- [FHKo17] Fachhochschule Köln (2011). Datenbanken Online Lexikon.
URL: http://wikis.gm.fh-koeln.de/wiki_db/Datenbanken

Literaturverzeichnis

- [GaUW08] Garcia-Molina, H., Ullman, J. D., Widom, J. (2008). Database Systems: The Complete Book. Pearson International.
- [GMPB+09] Ginsberg, J., Mohebbi, M. H., Patel, R. S., Brammer, L., Smolinski, M. S., & Brilliant, L. (2009). Detecting influenza epidemics using search engine query data. Nature, 457 (7232), 1012-1014.
- [HaKa06] Han, J., Kamber, M. (2006). Data mining: concepts and techniques. Morgan Kaufmann.
- [Harr15] Harrison, G. P. (2015). Next Generation Databases - NoSQL and Big Data. Apress.
- [JZFF11] Jannach, D., Zanker, M., Felfernig, A., & Friedrich, G. (2010). Recommender Systems: An Introduction. Cambridge University Press.
- [KeBM10] Kemper, H. G., Baars, H., Mehanna, W. (2010). Business Intelligence: Grundlagen und praktische Anwendungen. Vieweg.

Literaturverzeichnis

- [KeEi15] Kemper, A., Eickler, A. (2015). Datenbanksysteme. Oldenbourg.
- [KrGL05] Kraft, M., Götz, O., Liehr-Gobbers, K. (2005). Die Validierung von Strukturgleichungsmodellen mit Hilfe des Partial-Least-Squares-Ansatzes, in: Handbuch PLS-Pfadmodellierung. Schäffer Poeschel.
- [KuJo13] Kuhn, M., Johnson, K. (2013): Applied Predictive Modeling. Springer.
- [LKKV14] Lazer, D., Kennedy, R., King, G., Vespignani, A. (2014). The parable of Google Flu: traps in big data analysis. Science, 343 (6176), 1203-1205.
- [MBK+12] Mertens, P., Bodendorf, F., König, W., Schumann, M., Hess, T. (2012). Grundzüge der Wirtschaftsinformatik. Springer.
- [Muel15] Müller, E. (2015). Vorlesungsskript: Big Data Analytics. Hasso-Plattner-Institut Potsdam.
- [Naum15] Naumann, F. (2015). Vorlesungsskript: Datenbanksysteme II. Hasso-Plattner-Institut Potsdam.

Literaturverzeichnis

[O'Neil 6] O'Neil, C. (2016). Weapons of math destruction: How big data increases inequality and threatens democracy. Crown Books.

[PäPa 15] Pääkkönen, P., Pakkala, D. (2015). Reference architecture and classification of technologies, products and services for big data systems. Big Data Research, 2 (4), 166-186.

[Plat 13] Plattner, H. (2013). A Course in In-Memory Data Management. Springer.

[PoBe 13] Becker, L., Pousttchi, K. (2013). Requirements for personalized m-commerce - what drives consumers' use of social networks?. In: Journal of Electronic Commerce in Organizations 11 (2013) 4, S. 19-36.

[RaSS 15] Rahm, E., Saake, G., & Sattler, K. U. (2015). Verteiltes und paralleles Datenmanagement: von verteilten Datenbanken zu Big Data und Cloud. Springer.

[RePi 91] Reichwald, R., Picot, A. (1991). Informationswirtschaft, in: Heinen, E. (Hrsg.). Industriebetriebslehre: Entscheidungen im Industriebetrieb. Springer.

[Rupa 16] Ruparelia, N. B. (2016). Cloud Computing. MIT Press.

Literaturverzeichnis

- [RuQZ07] Rupp, C., Queins, S., Zengler, B. (2007). UML 2 glasklar: Praxiswissen für die UML-Modellierung. Carl Hanser.
- [SaSH11] Saake, G., Sattler, K. U., Heuer, A. (2011). Datenbanken: Implementierungstechniken. mitp.
- [SaSH13] Saake, G., Sattler, K. U., Heuer, A. (2013). Datenbanken: Konzepte und Sprachen. mitp.
- [Sche97] Scheer, A. W. (1997). ARIS—vom Geschäftsprozess zum Anwendungssystem. Springer.
- [StHa04] Stahlknecht, P., Hasenkamp, U. (2004). Einführung in die Wirtschaftsinformatik. Springer.
- [Sull15] Sullivan, D. (2015). NoSQL for mere mortals. Addison-Wesley Professional.
- [Thor13] Thor, A. (2013). Vorlesungsskript: Datenmanagement. Universität Leipzig.



Literaturverzeichnis

- [Vige16] Vigen, T. (2016). Spurious Correlations. Hachette Books.
- [Voss08] Vossen, G. (2008). Datenmodelle, Datenbanksprachen und Datenbank-Management-Systeme. Oldenbourg.
- [WeKa16] Wedel, M., Kannan, P. K. (2016). Marketing analytics for data-rich environments. Journal of Marketing, 80(6), 97-121.
- [Wies15] Wiese, L. (2015). Advanced Data Management: For SQL, NoSQL, Cloud and Distributed Databases. de Gruyter.
- [Zehn13] Zehnder, C. A. (2013). Informationssysteme und Datenbanken. Springer.
- [Zoss15] Zoss, A. (2015). Practical Data Visualization. Lecture at Duke University.