

Datenbanksysteme

Martin Schäler
martin.schaeler@kit.edu

Lehrstuhl für Systeme der Informationsverwaltung, Fakultät für Informatik



Photo by Beinecke Library

Vorlesungstermine

- Montags 24.04., 08.05. zum Übungstermin – Vorlesung.
- Eventuell noch ein, zwei weitere Montagstermine, anstelle von Donnerstagsterminen. Wird rechtzeitig bekanntgegeben.

Organisatorisches

- Sie können die Folien vorab im Ilias herunterladen.
 - Alles wesentliche auch im Ilias angekündigt
- Vorlesungsmitschnitt von 2015 (gleicher Inhalt wie dieses Jahr):
<https://videos.informatik.kit.edu>

Erwartungen unsererseits

- Nacharbeiten der Vorlesung
 - in der gleichen Woche wie die jeweilige Sitzung.
- Nachdenken über ‚mögliche Prüfungsfragen‘.
Können Sie sie auch beantworten?
- Bearbeiten der Übungsaufgaben
(auch derer, die Sie nicht abgeben müssen).
- Fragen während der Vorlesung sind erwünscht.

Prüfung

Für Bachelor-/Master-Studierende:

- Klausur am Semesterende
(Termin voraussichtlich Ende Juli/Anfang August 2017;
Nachtermin voraussichtlich
in der zweiten Februarhälfte 2018)
- Eine Übungsaufgabe zum Abgeben:
Notenbonus für diesjährige Klausur
bei erfolgreicher Bearbeitung der Aufgabe.

Organisatorisches, Fortsetzung

- Sie müssen sich über das Studierendenportal (campus.studium.kit.edu) für die Prüfung anmelden. Wie für andere Prüfungen auch.
- Ggf. Abmeldung innerhalb der üblichen Fristen auch dort.

Literatur

- Andreas Heuer, Kai-Uwe Sattler, Gunter Saake:
Datenbanken – Konzepte und Sprachen
mitp-Verlag, Bonn
- Alfons Kemper, André Eickler:
Datenbanksysteme. Eine Einführung
Oldenbourg Verlag
- Hector Garcia-Molina, Jeffrey D. Ullman, Jennifer Widom:
Database Systems: The Complete Book
Prentice Hall, 2008
- Gerhard Weikum, Gottfried Vossen:
Transactional Information Systems
Morgan Kaufmann, 2002
- Eric Redmond, Jim R. Wilson: *Seven Databases in Seven Weeks:
A Guide to Modern Databases and the NoSQL Movement*
The Pragmatic Programmers, 2012.

Datenbanksysteme

Kapitel 1: Grundlegende Konzepte und Architekturen

Lehrstuhl für Systeme der Informationsverwaltung, Fakultät für Informatik



Photo by Beinecke Library

Motivation

- Unser Thema: Verwaltung großer Datenmengen und Entwicklung entsprechender Anwendungen.
- Warum Datenbanken?
- *Application-Development Complexity.*
 - 50% weniger Aufwand bei Anwendungsentwicklung mit Datenbanken, gesamthaft betrachtet, bei überschaubaren Projekten.
 - Größere Differenz bei größeren Projekten.
⇒ Verwendung von Datenbank-Technologie ermöglicht neue Anwendungen; ihre Entwicklung ohne Datenbank-Technologie wäre zu komplex.

Einleitung

Welt
ohne DB

Datenbanken
Übersicht

Terminologie

Querying

DB –
Merkmale

Anwend.

Vergleich mit Bekanntem

- Informatik – zwei wichtige Aspekte:
 - Erkennen von Regelmäßigkeiten/Mustern bei der Anwendungsentwicklung.
Ausfaktorisieren entsprechender Funktionalität.
Beispiele:
 - Grafik-Pakete,
 - Kommunikation, Middleware,
 - Wissenschaftliches Rechnen,
 - SAP.
 - Schaffung von Konzepten für die Programmierung auf abstrakterer Ebene.
Z. B. Objekt- oder Aspektorientierung
sowie stets modernere Entwicklungsumgebungen.
- Datenbanken –
dasselbe für Entwicklung von Anwendungen zur Verwaltung großer Datenmengen.

Einleitung

Welt
ohne DB

Datenbanken
Übersicht

Terminologie

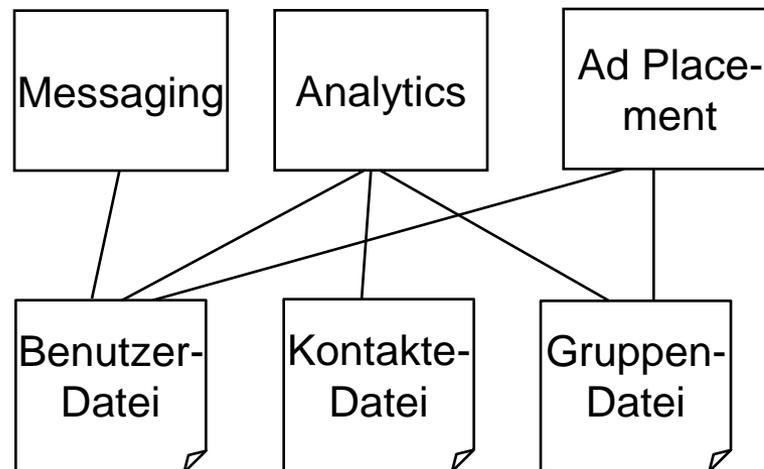
Querying

DB –
Merkmale

Anwend.

Situation ohne Datenbanken (1)

- Zugriff auf Daten, in Dateien abgelegt.
- Funktionalität hierfür Teil der Anwendungen.



Einleitung

Welt
ohne DB

Datenbanken
Übersicht

Terminologie

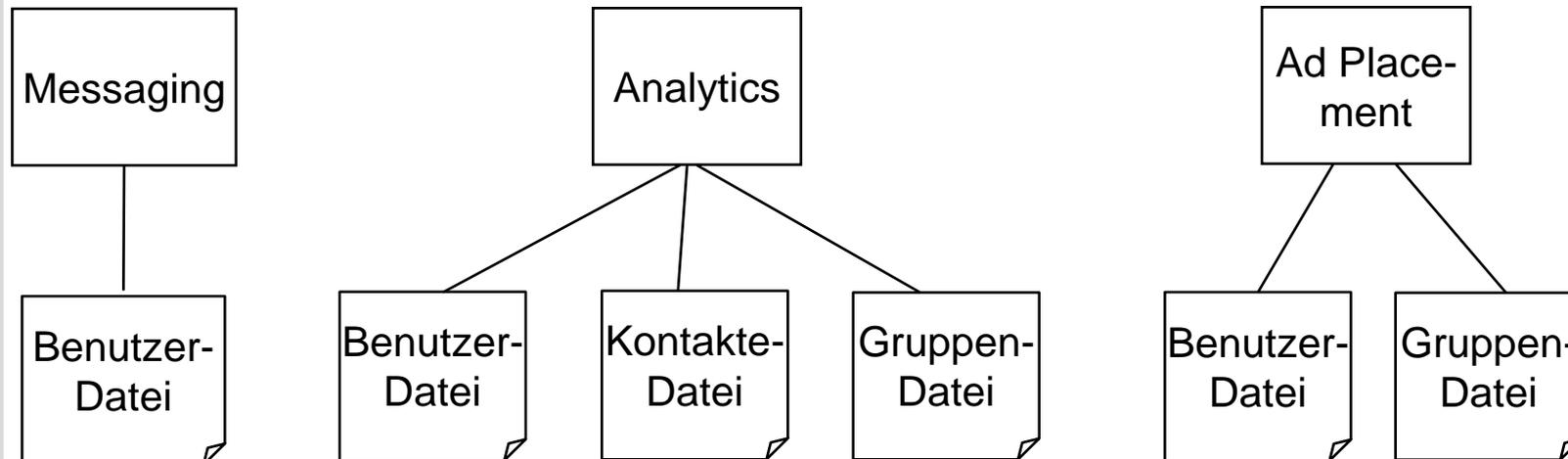
Querying

DB –
Merkmale

Anwend.

Situation ohne Datenbanken (2)

■ Redundanz.



Einleitung

Welt
ohne DB

Datenbanken
Übersicht

Terminologie

Querying

DB –
Merkmale

Anwend.

Ohne Datenbanken: Datenredundanz (1)

- Daten sind *redundant*: Mehrfach gespeichert;
Probleme: Verschwendung von Speicherplatz,
„Vergessen“ von Änderungen, *Inkonsistenzen*;
keine zentrale, „genormte“ Datenhaltung.
- Redundanz lässt sich zwar auch ohne Datenbanken vermeiden, Datenbanken unterstützen dies aber gut.
- *Auch Funktionalität ist redundant*,
d. h. kommt in den unterschiedlichen Anwendungen wiederholt vor.

Einleitung

Welt
ohne DB

Datenbanken
Übersicht

Terminologie

Querying

DB –
Merkmale

Anwend.

Free
cash withdrawals

We won't charge you for cash withdrawals using your debit card



HSBC



- Top up mobile phones
- Check Salaries
- Check card and balance on pin
- Make payments
- Pay HSBC credit cards

Clydesdale Bank

VISA  

Free Cash Withdrawals
Clydesdale Bank will not charge you for using the machine

Clydesdale Bank



Mobile Phone Top Up

RBS
The Royal Bank of Scotland

CASHLINE



Weitere Probleme ohne Datenbanken (1)

- Transaktionseigenschaften – insbesondere *Atomarität* und *Isolation*.

- Atomarität

- Beispiel, „Bank-Szenario“:

Nummer	Inhaber	Stand
	Klemens	5000
	Gunter	200

- Überweisung – zwei Elementaroperationen.
 - Abbuchung(Klemens, 500),
 - Einzahlung(Gunter, 500).
 - *Atomarität* – Programm wird entweder ganz oder gar nicht ausgeführt.

Einleitung

Welt
ohne DB

Datenbanken
Übersicht

Terminologie

Querying

DB –
Merkmale

Anwend.

Weitere Probleme ohne Datenbanken (2)

■ Isolation

- Beispiel, „Bank-Szenario“:

<u>Nummer</u>	<u>Inhaber</u>	<u>Stand</u>
	Klemens	5000
	Gunter	200

- Überweisung – zwei Elementaroperationen.
 - Abbuchung(Klemens, 500),
 - Einzahlung(Gunter, 500).
- *Isolation* – keine inkonsistenten Zwischenzustände werden sichtbar.
- Transaktionen.
Programm oder Folge von Kommandos,
die in Interpreter eingegeben werden.

Einleitung

Welt
ohne DB

Datenbanken
Übersicht

Terminologie

Querying

DB –
Merkmale

Anwend.

Weitere Probleme ohne Datenbanken (3)

- Atomarität – hat mit Mehrbenutzerbetrieb nichts zu tun.
- Atomarität zielt auf den Fehlerfall. Ist bereits bei Ausführung eines Programms bedeutsam.
- Isolation hingegen:
 - Nur wichtig, wenn mehrere Transaktionen.
 - Selbst in Systemen bedeutsam, in denen es keine Fehler gibt.

Einleitung

Welt
ohne DB

Datenbanken
Übersicht

Terminologie

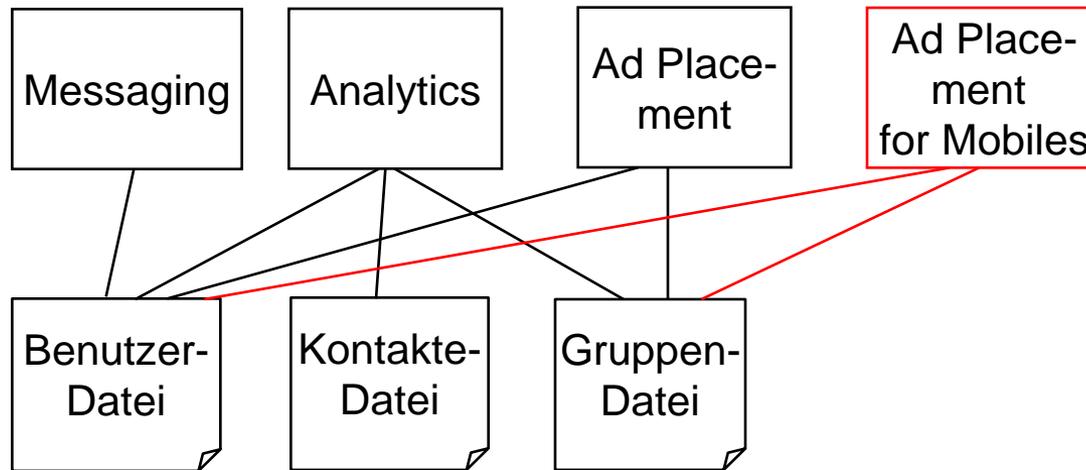
Querying

DB –
Merkmale

Anwend.

Weitere Probleme ohne Datenbanken (4)

- Mehrere Benutzer oder Anwendungen können i. Allg. nicht parallel auf den gleichen Daten arbeiten („Nebenläufigkeit“, *Concurrency*), ohne sich zu stören.



Einleitung

Welt
ohne DB

Datenbanken
Übersicht

Terminologie

Querying

DB –
Merkmale

Anwend.

- Es wäre extrem aufwendig, Lösungen hierfür selbst entwickeln und realisieren zu müssen.

Weitere Probleme ohne Datenbanken (5)

- Keine *physische Datenunabhängigkeit*:

- ‚Naives‘ file-basiertes Vorgehen bei Programmierung/

Benutzung von Anwendungen:

Man muss interne (physische) Repräsentation der Daten kennen,

z. B. bei Speicherung als Tabelle

Reihenfolge der Zeilen und Spalten.

Einleitung

Welt
ohne DB

Datenbanken
Übersicht

Terminologie

Querying

DB –
Merkmale

Anwend.

Weitere Probleme ohne Datenbanken (6)

- Keine *physische Datenunabhängigkeit* (Forts.):
 - Beispiel „Benutzerdaten eines sozialen Netzwerks“ (SNUSER):

NAME	VORNAME	STRASSE	ALTER
Böhm	Klemens	Fasanengarten	32
Buchmann	Erik	Breiter Weg	26
Schäler	Martin	Goethestrasse	25
Saake	Gunter	Waldweg	43

VORNAME	NAME	STRASSE	ALTER
Erik	Buchmann	Breiter Weg	26
Gunter	Saake	Waldweg	43
Klemens	Böhm	Fasanengarten	32
Martin	Schäler	Goethestrasse	25

Einleitung

Welt
ohne DB

Datenbanken
Übersicht

Terminologie

Querying

DB –
Merkmale

Anwend.

Reihenfolge der Zeilen bedeutsam. Dto. Reihenfolge der Spalten.

Weitere Probleme ohne Datenbanken (7)

- Keine *physische Datenunabhängigkeit* (Forts.):
 - Physische Datenunabhängigkeit – Verstecken dieser physischen Eigenschaften vor dem Anwendungsentwickler.
 - Datenbanken geben uns physische Datenunabhängigkeit (ein herausragendes Merkmal).
- Beispiel, fortgesetzt – SQL als Anfragesprache:

```
SELECT VORNAME, NAME
FROM SNUSER
WHERE ALTER < 30
```

Einleitung

Welt
ohne DB

Datenbanken
Übersicht

Terminologie

Querying

DB –
Merkmale

Anwend.

Weitere Probleme ohne Datenbanken (8)

- Datenschutz und Datensicherheit sind nicht gewährleistet.
 - Datenschutz (kein unbefugter Zugriff),
 - Datensicherheit (kein ungewollter Datenverlust, insbesondere bei Defekten).

Einleitung

Welt
ohne DB

Datenbanken
Übersicht

Terminologie

Querying

DB –
Merkmale

Anwend.

Von Dateien zu Datenbanken

- Zugriff auf Daten, in Files abgelegt.
- Funktionalität hierfür Teil der Anwendungen
(Berücksichtigung der physischen Ebene, Nebenläufigkeit, Datenschutz, Konsistenz).
- Datenbanken:
Ausfaktorisieren dieser Funktionalität.

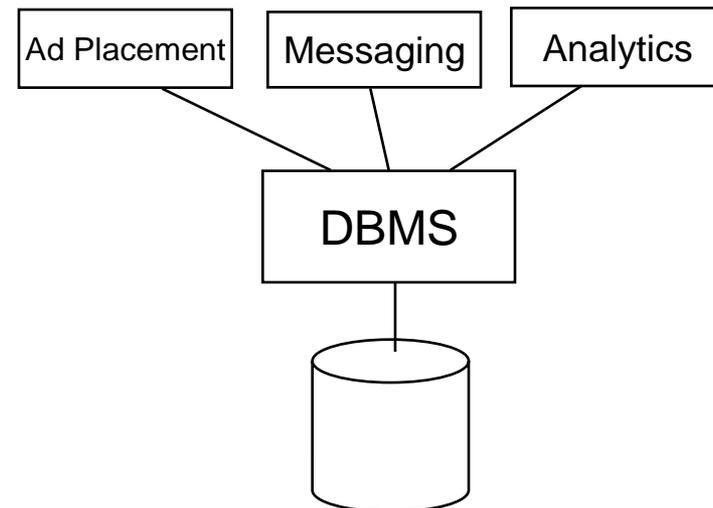
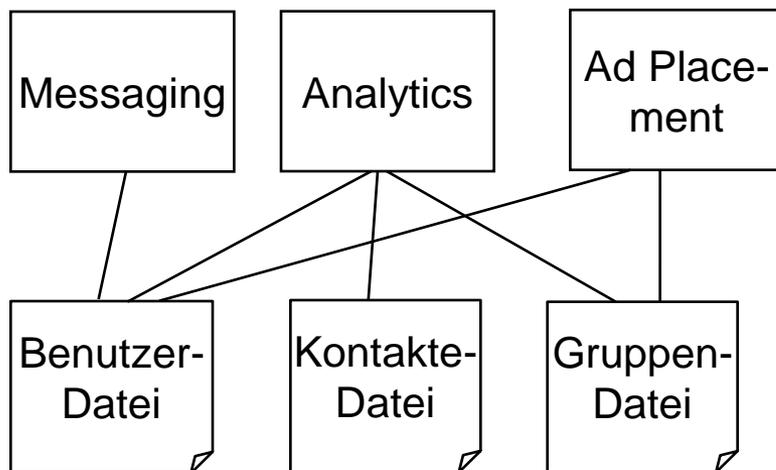
Einleitung

Welt
ohne DB

Datenbanken
Übersicht

Terminologie

Querying
DB –
Merkmale
Anwend.



Mögliche Prüfungsfragen

- Was für Probleme bringt Datenhaltung ohne Datenbanken mit sich?
- Wie profitiert ein Anwendungsentwickler von der Verwendung von Datenbank-Technologie?
- Erklären Sie die folgenden Begriffe:
 - Redundanz, Inkonsistenz,
 - Atomarität, Isolation, Transaktion,
 - physische Datenunabhängigkeit (weitere Erläuterungen hierzu folgen).

Einleitung

Welt
ohne DB

Datenbanken
Übersicht

Terminologie

Querying

DB –
Merkmale

Anwend.

Datenbanktechnologie – Vielfalt (1)

- Relationale Datenbanken (RDBMSen)
 - Marktführende Produkte:
z.B. Oracle, MS SQL-Server, IBM DB2
 - Weiterentwicklungen, sowohl eher akademisch (z.B. NF2)
als auch kommerziell (z.B. objekt-relationale DBMSen).
[inzwischen weitgehend tot]
- NoSQL-Datenbanken
 - Key-Value Stores,
 - spaltenorientierte Datenbanken à la HBase,
 - dokumentorientierte Datenbanken mit dynamischem Schema,
z. B. MongoDB; vormals XML-Datenbanken,
 - Graph-Datenbanken.
- Werkzeuge für Datenanalyse
(orthogonal zu den o. g. Technologien).

Einleitung

Welt
ohne DB

Datenbanken
Übersicht

Terminologie

Querying

DB –
Merkmale

Anwend.

Datenbanktechnologie – Vielfalt (2)

- Je nach Anforderungen ist unterschiedliche Technologie besser geeignet.
- Im Folgenden kurze Vorstellung klassischer DB-Features am Beispiel von RDBMSen.

Einleitung

Welt
ohne DB

Datenbanken
Übersicht

Terminologie

Querying

DB –
Merkmale

Anwend.

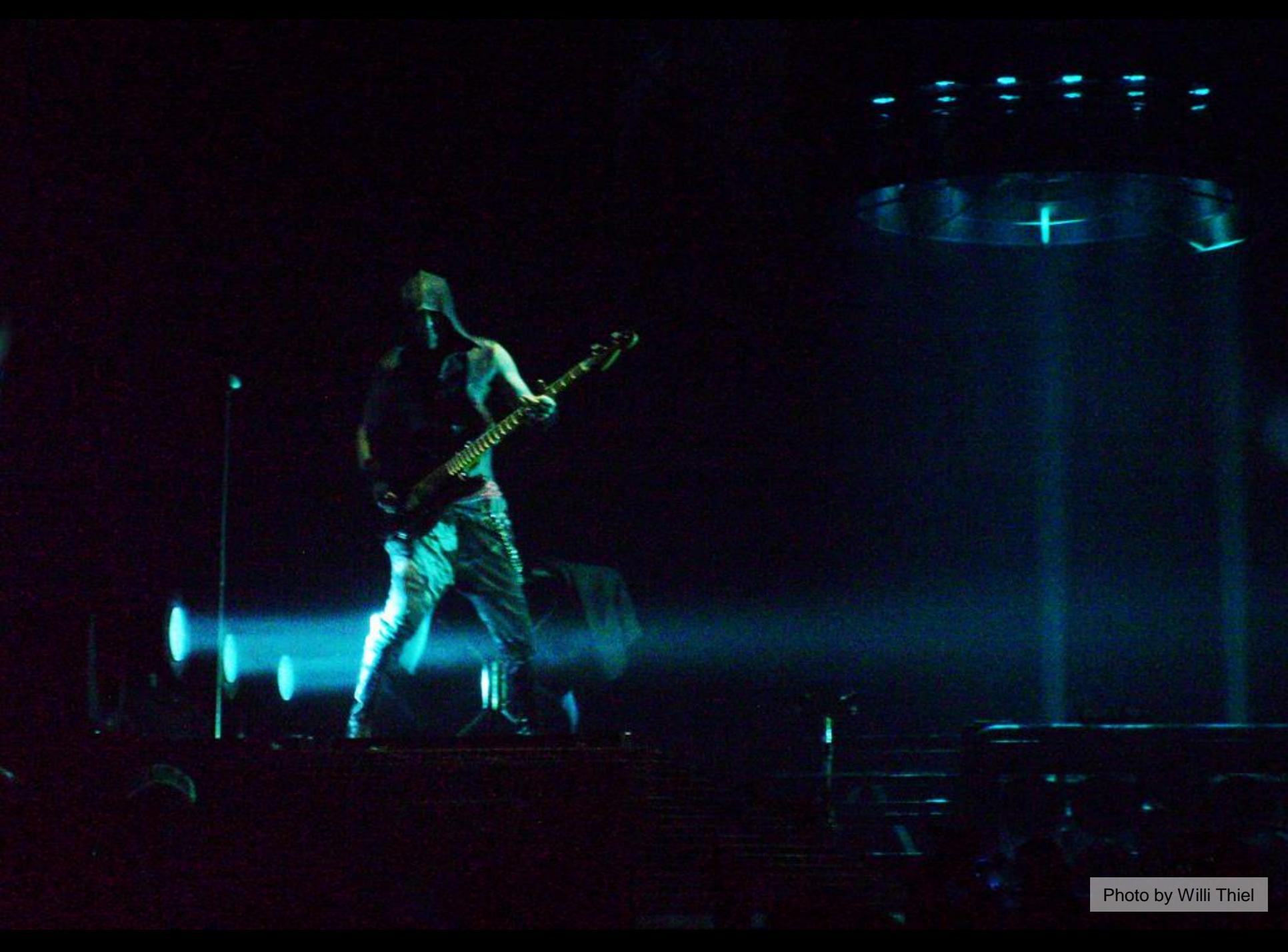


Photo by Willi Thiel

Konzeptuelle Ebene: Relationenmodell (1)

- Konzeptuell ist RDBMS Menge von Tabellen.

Künstler	KID	NAME	LAND	JAHR
	1012	Neil Young	Kanada	1945
	1014	Rammstein	Deutschland	1994
	1015	The Ramones	USA	1974
	1016	Eric Fish	Deutschland	1969

Titel

TITLE ID	NAME	ART	GRÖSSE	KID
102	Neil Young – Heart of Gold	mp3	2.920kb	1012
103	Rammstein – Ich liebe Neil Young	wma	4.234kb	1014
104	Neil Young – Old Man	mp3	3.161kb	1012
105	Neil Young – Four Strong Winds	wma	5.125kb	1012

Einleitung

Welt
ohne DB

Datenbanken
Übersicht

Terminologie

Querying

DB –
Merkmale

Anwend.

Konzeptuelle Ebene: Relationenmodell (2)

- Tabellen = „Relationen“
- Relationen sind Mengen von Tupeln.
- **Fett** geschriebene Zeilen: Relationenschema
- Weitere Einträge in der Tabelle: *Relation*
- Eine Zeile der Tabelle: *Tupel*
- Eine Spaltenüberschrift: *Attribut*
- Name der Tabelle: *Relationenname*

Einleitung

Welt
ohne DB

Datenbanken
Übersicht

Terminologie

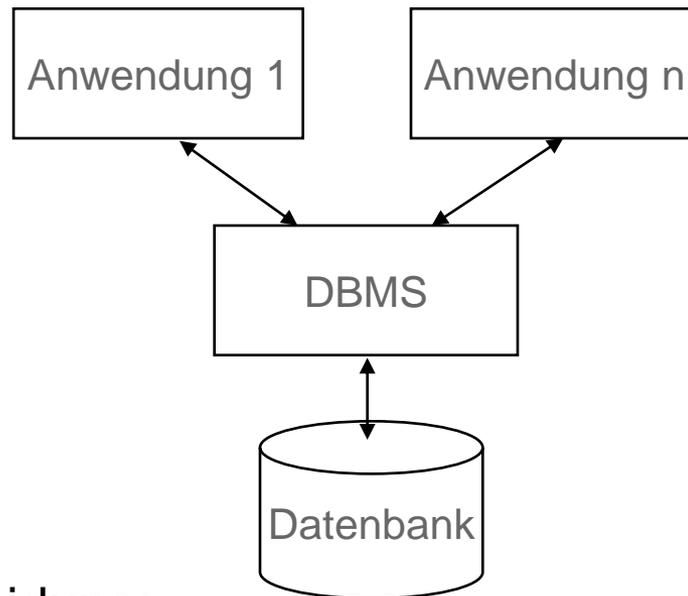
Querying

DB –
Merkmale

Anwend.

DBMS vs. DBS

- DBMS: Datenbank-Management-System
- DBS: Datenbanksystem (DBMS + Datenbank)



- Unterscheidung:

- DBMS ist Software zur Datenverwaltung, Datenbank sind die eigentlichen Daten.
- Ein DBMS, aber mehrere (viele) Datenbanken.

Einleitung

Welt
ohne DB

Datenbanken
Übersicht

Terminologie

Querying

DB –
Merkmale

Anwend.

Integritätsbedingungen am Beispiel

- Konzeptuell ist RDBMS Menge von Tabellen.

Künstler	KID	NAME	LAND	JAHR
	1012	Neil Young	Kanada	1945
	1014	Rammstein	Deutschland	1994
	1015	The Ramones	USA	1974
	1016	Eric Fish	Deutschland	1969

Titel

TITLE ID	NAME	ART	GRÖSSE	KID
102	Neil Young – Heart of Gold	mp3	2.920kb	1012
103	Rammstein – Ich liebe Neil Young	wma	4.234kb	1014
104	Neil Young – Old Man	mp3	3.161kb	1012
105	Neil Young – Four Strong Winds	wma	5.125kb	1012

Einleitung

Welt
ohne DB

Datenbanken
Übersicht

Terminologie

Querying

DB –
Merkmale

Anwend.

Integritätsbedingungen

- *Relationenschema*
+ *lokale Integritätsbedingungen:*
 - KID ist *Schlüssel* für **Künstler**,
 - d. h. KID darf nicht doppelt vergeben werden.
- *Datenbankschema*
ist Menge von Relationenschemata
+ *globale Integritätsbedingungen:*
 - KID in Titel
ist *Fremdschlüssel bezüglich Künstler*,
 - d. h.: KID taucht
in einem anderen Relationenschema als Schlüssel auf.

Einleitung

Welt
ohne DB

Datenbanken
Übersicht

Terminologie

Querying

DB –
Merkmale

Anwend.

Mögliche Prüfungsfragen

- Was versteht man unter Datenbank-Schema?
- Erklären Sie den Unterschied zwischen ‚Datenbank‘ und ‚DBMS‘.
- Was für Arten von Integritätsbedingungen, die DBMS unterstützen, kennen Sie? (Kommt noch einmal ausführlicher.)

Einleitung

Welt
ohne DB

Datenbanken
Übersicht

Terminologie

Querying

DB –
Merkmale

Anwend.

Sprachen und Sichten (1)

- Wie greife ich auf Informationen in meiner Datenbank zu?
- SQL als Standard-Anfragesprache für RDBMS.
- *Anfragen, Queries* – dasselbe.
- Query-Ergebnis – auch wieder Relation.

Einleitung

Welt
ohne DB

Datenbanken
Übersicht

Terminologie

Querying

DB –
Merkmale

Anwend.

✓ Zeige Datensätze 0 - 3 (4 insgesamt, Die Abfrage dauerte 0.0001 Sekunden)

```
SELECT * FROM `Titel`
```

Messen [Inline] [Bearbeiten] [SQL erklären] [PHP-Code erzeugen] [Aktualisieren]

Anzahl der Datensätze: 25

Nach Schlüssel sortieren: keine

+ Optionen

	TITLEID	NAME	ART	GROESSE	KID
<input type="checkbox"/> Bearbeiten Kopieren Löschen	102	Neil Young - Hart of Gold	mp3	2.920kb	1012
<input type="checkbox"/> Bearbeiten Kopieren Löschen	103	Rammstein - Ich liebe Neil Young	wma	4.234kb	1014
<input type="checkbox"/> Bearbeiten Kopieren Löschen	104	Neil Young - Old Man	mp3	3.161kb	1012
<input type="checkbox"/> Bearbeiten Kopieren Löschen	105	Neil Young - Four Strong Winds	wma	5.125kb	1012

Alle auswählen markierte: Bearbeiten Löschen Exportieren

Anzahl der Datensätze: 25

✓ Zeige Datensätze 0 - 1 (2 insgesamt, Die Abfrage dauerte 0.0003 Sekunden)

```
SELECT * FROM `Titel` where ART = 'mp3'
```

Messen [Inline] [Bearbeiten] [SQL erklären] [PHP-Code erzeugen] [Aktualisieren]

Anzahl der Datensätze: 25

Nach Schlüssel sortieren: keine

+ Optionen

			TITLEID	NAME	ART	GROESSE	KID	
<input type="checkbox"/>	Bearbeiten	Kopieren	Löschen	102	Neil Young - Hart of Gold	mp3	2.920kb	1012
<input type="checkbox"/>	Bearbeiten	Kopieren	Löschen	104	Neil Young - Old Man	mp3	3.161kb	1012

Alle auswählen *markierte:* Bearbeiten Löschen Exportieren

Anzahl der Datensätze: 25

Operationen für das Abfrageergebnis

Sprachen und Sichten (3)

- Definition von Benutzersichten:
Häufig vorkommende Datenbankabfragen (Queries) können unter einem „Sichtnamen“ als „virtuelle“ Tabelle gespeichert werden.
- *Sicht* = Query (+ Name),
- englisch: View.

Einleitung

Welt
ohne DB

Datenbanken
Übersicht

Terminologie

Querying

DB –
Merkmale

Anwend.

Definition von Sichten in SQL – Beispiel

- Zugrundeliegende Relation:
Künstler(KID, NAME, LAND, JAHR)
- **create view** CArtist **as**
select NAME, JAHR
from Künstler
where LAND == "Kanada"
- View dann im Prinzip verwendbar
wie ‚normale‘ Relation, z. B.:
select * **from** CArtist **where** JAHR < 2000
- ‚Rückkopplung‘ zur zugrundeliegenden Relation. i
- Sicht ist virtuelles Konzept.

Einleitung

Welt
ohne DB

Datenbanken
Übersicht

Terminologie

Querying

DB –
Merkmale

Anwend.

Sprachen und Sichten (4)

- Sichten werden gern kombiniert mit Datenschutz-Mechanismen.
- Unterschiedliche Benutzer sehen unterschiedlichen Ausschnitt der Datenbank.
- Beispiel – Universität (Studierende, Dozenten, Prüfungsamt, Dekanat etc.):
 - Datenschutz,
 - Übersichtlichkeit,
 - organisatorische Gründe (Ablenkung, Einmischung).

Einleitung

Welt
ohne DB

Datenbanken
Übersicht

Terminologie

Querying

DB –
Merkmale

Anwend.

Anfrageauswertung am Beispiel

- Konzeptuell ist RDBMS Menge von Tabellen.

Künstler	KID	NAME	LAND	JAHR
	1012	Neil Young	Kanada	1945
	1014	Rammstein	Deutschland	1994
	1015	The Ramones	USA	1974
	1016	Eric Fish	Deutschland	1969

Titel

TITLE ID	NAME	ART	GRÖSSE	KID
102	Neil Young – Heart of Gold	mp3	2.920kb	1012
103	Rammstein – Ich liebe Neil Young	wma	4.234kb	1014
104	Neil Young – Old Man	mp3	3.161kb	1012
105	Neil Young – Four Strong Winds	wma	5.125kb	1012

Einleitung

Welt
ohne DB

Datenbanken
Übersicht

Terminologie

Querying

DB –
Merkmale

Anwend.

Anfrageoperationen (1)

- SELEKTION: Zeilen (Tupel) auswählen.

$\sigma_{KID=1012}(\text{Titel})$

TITLE ID	NAME	ART	GRÖSSE	KID
102	Neil Young – Heart of Gold	mp3	2.920kb	1012
104	Neil Young – Old Man	mp3	3.161kb	1012
105	Neil Young – Four Strong Winds	wma	5.125kb	1012

Einleitung

Welt
ohne DB

Datenbanken
Übersicht

Terminologie

Querying

DB –
Merkmale

Anwend.

- PROJEKTION: Spalten (Attribute) auswählen.

$\pi_{KID, NAME}(\text{Künstler})$

KID	NAME
1012	Neil Young
1014	Rammstein
1015	The Ramones
1016	Eric Fish

Anfrageoperationen (2)

- Beispiel für komplexen Algebraausdruck:

$$\pi_{\text{NAME, ART}}(\sigma_{\text{KID}=1012}(\text{Titel}))$$

Ausgangsrelation:

TITLE ID	NAME	ART	GRÖSSE	KID
102	Neil Young – Heart of Gold	mp3	2.920kb	1012
103	Rammstein – Ich liebe Neil Young	wma	4.234kb	1014
104	Neil Young – Old Man	mp3	3.161kb	1012
105	Neil Young – Four Strong Winds	wma	5.125kb	1012

Einleitung

Welt

ohne DB

Datenbanken

Übersicht

Terminologie

Querying

DB –

Merkmale

Anwend.

Ergebnis:

NAME	ART
Neil Young – Heart of Gold	mp3
Neil Young – Old Man	mp3
Neil Young – Four Strong Winds	wma

Anfrageoperationen (3)

- Weitere Operationen: Verbund (Join), Vereinigung, Differenz, Durchschnitt, Umbenennung.
- Alle Operationen beliebig kombinierbar („Algebra“, *Query-Algebra*).

Einleitung

Welt
ohne DB

Datenbanken
Übersicht

Terminologie

Querying

DB –
Merkmale

Anwend.

Optimierer: Algebraische Optimierung

■ Allgemeine Regel:

1. $\sigma_{A1 = Konst}(\sigma_{A2 = Konst}(REL))$

2. $\sigma_{A2 = Konst}(\sigma_{A1 = Konst}(REL))$

sind äquivalent.

■ Beispiel:

1. $\sigma_{Vorname = 'Klemens'}(\sigma_{Wohnort = 'KA'}(SNUSER))$

2. $\sigma_{Wohnort = 'KA'}(\sigma_{Vorname = 'Klemens'}(SNUSER))$

■ Zweite Variante

bei ‚natürlicher‘ Verteilung der Daten überlegen.

(Verarbeitungsmodell

– ein Operator nach dem anderen,

Ergebnis zwischenspeichern.)

Einleitung

Welt
ohne DB

Datenbanken
Übersicht

Terminologie

Querying

DB –
Merkmale

Anwend.

Optimierer

- Problem: Finde einen Relationenalgebra-Ausdruck, der äquivalent ist („das gleiche Ergebnis liefert“) zum gegebenen, aber effizienter auszuwerten ist.
- Aufgabe des *Anfrage-Optimierers*.

Einleitung

Welt
ohne DB

Datenbanken
Übersicht

Terminologie

Querying

DB –
Merkmale

Anwend.

Optimierer & Physische Datenunabhängigkeit

- Anfragen sind *deklarativ*, Anwender sagt nur, welches Ergebnis, nicht, wie es ermittelt werden soll.
- Beispiel:
select *
from SNUSER
where VORNAME = 'Klemens'
and WOHNORT = 'KA'
- Datenbank 1:
1 Tupel mit VORNAME = 'Klemens',
1000 Tupel mit WOHNORT = 'KA'.
- Datenbank 2: umgekehrt.

Einleitung

Welt
ohne DB

Datenbanken
Übersicht

Terminologie

Querying

DB –
Merkmale

Anwend.

Optimierer & Physische Datenunabhängigkeit

- Physische Datenunabhängigkeit – DBMS stellt sicher,
 - dass Anfrage weiterhin gut funktioniert, auch wenn physische Darstellung der Daten sich geändert hat.
 - dass Anfrage in unterschiedlichen Datenbanken (gleiches Schema, aber unterschiedliche Häufigkeiten der Daten) funktioniert.
- Erlaubt höhere Komplexität bei Anwendungsentwicklung.

Einleitung

Welt
ohne DB

Datenbanken
Übersicht

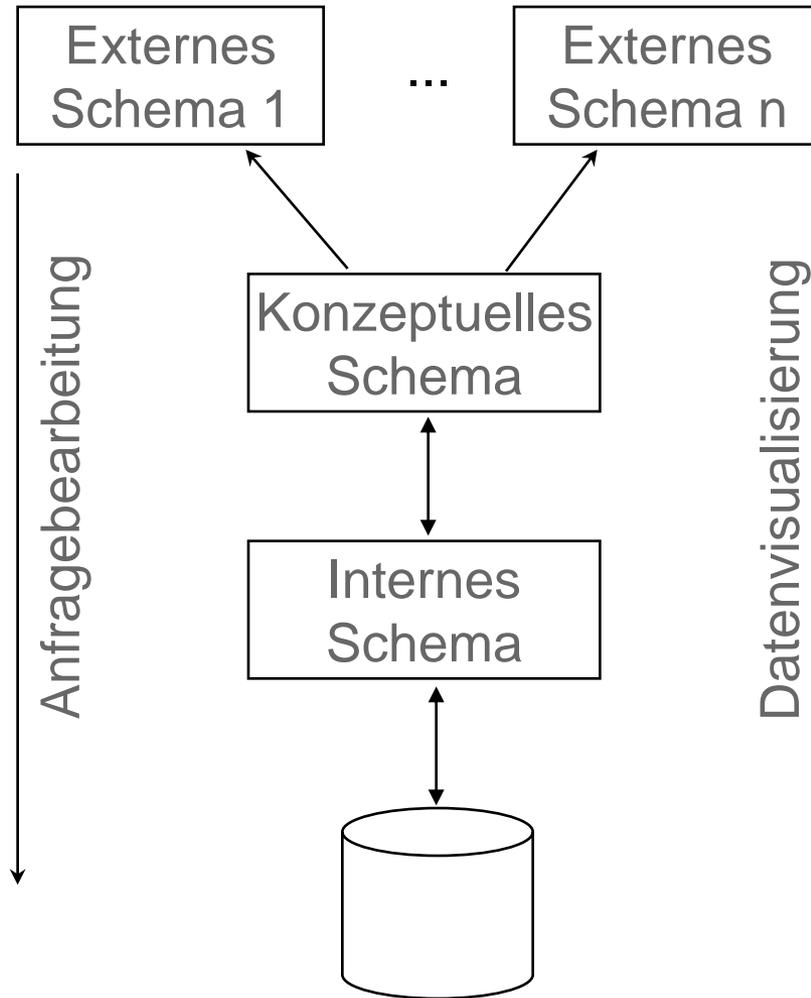
Terminologie

Querying

DB –
Merkmale

Anwend.

3-Ebenen Architektur (1)



Einleitung

Welt
ohne DB

Datenbanken
Übersicht

Terminologie

Querying

DB –
Merkmale

Anwend.

3-Ebenen Architektur (2)

- Konzeptionelles Schema
 - Was ist mein Diskursbereich?
Welche Entitäten der realen Welt interessieren uns?
 - Z. B. bei Studierenden die Noten,
aber nicht die Hobbies oder Äußerlichkeiten.
- Internes Schema – physische Repräsentation der Daten.
- Vorgehen beim Entwurf:
Erst konzeptuelles Schema, dann internes.

Einleitung

Welt
ohne DB

Datenbanken
Übersicht

Terminologie

Querying

DB –
Merkmale

Anwend.

3-Ebenen Architektur (3)

■ Warum hat man externe Sichten?

Unterschiedlichen Ausschnitt der Daten für unterschiedliche Benutzer (Studierende, Professor, Dekan, Raumverwaltung, Prüfungsamt)

- Datenschutz,
- Übersichtlichkeit,
- organisatorische Gründe (Ablenkung, Einmischung),
- Verstecken von Änderungen am konzeptionellen Schema
 - Benutzer wird nicht irritiert,
 - Anwendungsprogramme laufen weiterhin.

Logische Datenunabhängigkeit.

Einleitung

Welt
ohne DB

Datenbanken
Übersicht

Terminologie

Querying

DB –
Merkmale

Anwend.

Zusammenfassung bis hierhin

- Grundprinzip moderner Datenbanksysteme
 - *3-Ebenen-Architektur*
(physische und logische Datenunabhängigkeit)
 - *Trennung zwischen Schema*
(etwa Tabellenstruktur) *und Instanz* (etwa Tabelleninhalt).

Einleitung

Welt
ohne DB

Datenbanken
Übersicht

Terminologie

Querying

DB –
Merkmale

Anwend.

Prinzipien

- Angelehnt an *9 Codd'sche Regeln*.
(Eigenschaften, die Datenbank haben muss, um als relational zu gelten.)

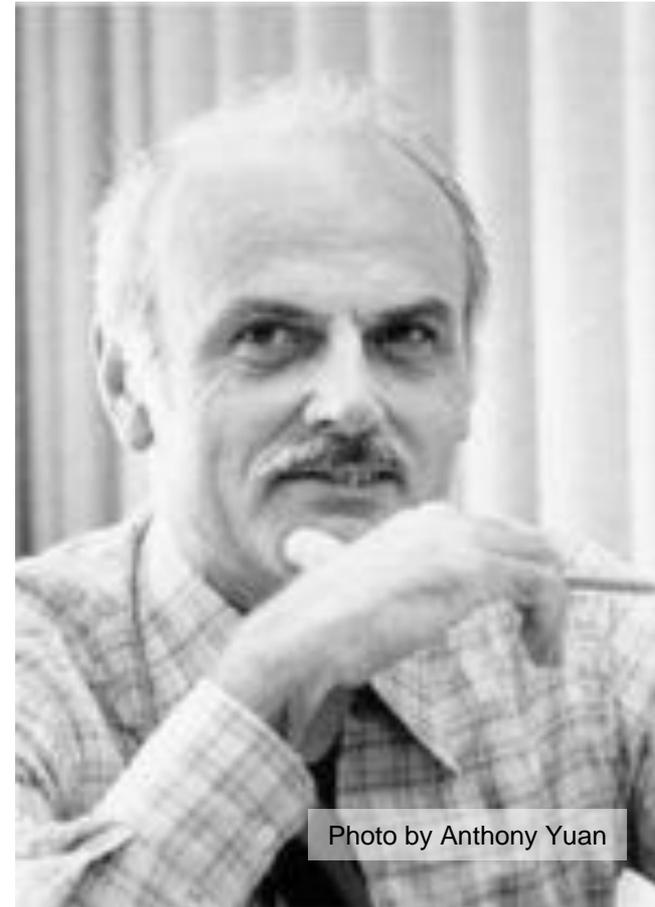


Photo by Anthony Yuan

Einleitung
Welt
ohne DB
Datenbanken
Übersicht
Terminologie
Querying
DB –
Merkmale
Anwend.

Die neun Codd'schen Regeln

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 unautorisierter Zugriffe,
7. Transaktionen:
mehrere DB-Operationen als Funktionseinheit (Atomarität),
8. Synchronisation:
parallele Transaktionen koordinieren (Isolation),
9. Datensicherung:
Wiederherstellung von Daten nach Systemfehlern.

Einleitung

Welt
ohne DB

Datenbanken
Übersicht

Terminologie

Querying

DB –
Merkmale

Anwend.

Mögliche Prüfungsfragen

- Was ist eine Sicht?
- Was ist die relationale Algebra?
Wozu braucht man sie?
- Geben Sie Beispiele für Algebra-Ausdrücke,
die nicht identisch, aber äquivalent sind?
- Was leistet der Anfrageoptimierer einer Datenbank?
- Erklären Sie: Drei-Ebenen Architektur,
physische Datenunabhängigkeit,
logische Datenunabhängigkeit.

Einleitung

Welt
ohne DB

Datenbanken
Übersicht

Terminologie

Querying

DB –
Merkmale

Anwend.

Funktionale vs. nichtfunktionale Anforderungen (1)

- Codd'sche Regeln sind strengste mir bekannte Definition von ‚Datenbank‘.
- Codd'sche Regeln sind funktionaler Natur.
- Nichtfunktionale Anforderungen hingegen:
 - Wie schnell/zuverlässig wird bestimmter Dienst unter bestimmten Umständen erbracht?
 - I. Allg. auch Abstufungen bei Ergebnisqualität vorstellbar/spezifizierbar.
(Z. B. im Prinzip „Jedes zehntausendste Anfrageergebnis muss nicht vollständig sein.“)

Einleitung

Welt
ohne DB

Datenbanken
Übersicht

Terminologie

Querying

DB –
Merkmale

Anwend.

Funktionale vs. nichtfunktionale Anforderungen (2)

■ *Service-Level Agreements:*

Vereinbarung zwischen Client und Server,
die Ausführung des Dienstes betreffend.

- Client-seitig z. B.:
Erwartete Rate von Dienstaufrufen,
- Server-seitig z. B.: Erwartete Latenzzeit
unter diesen Bedingungen.

Beispiel: “Antwort innerhalb von 300 ms für 99,9% der Aufrufe
bei einer Last von 500 Zugriffen pro Sekunde.”

Einleitung

Welt
ohne DB

Datenbanken
Übersicht

Terminologie

Querying

DB –
Merkmale

Anwend.

Aspekte nichtfunktionaler Anforderungen

- Kurze Antwortzeiten,
- Zuverlässigkeit,
- Effizienz (d. h. wie hoch sind die Kosten der Erbringung eines Dienstes, z. B. in Geld oder in Energiekosten),
- Skalierbarkeit, z. B. bezogen auf
 - Anfragen und Änderungsoperationen,
 - Datenvolumen.

Einleitung

Welt
ohne DB

Datenbanken
Übersicht

Terminologie

Querying

DB –
Merkmale

Anwend.



Photo by Scottish Government

Beispiel: Datenhaltung bei Amazon – Anforderungen (1)

- Krasse Anforderungen bezüglich Antwortzeiten, Zuverlässigkeit und Skalierbarkeit bezogen auf *Anfragevolumen*. Mit herkömmlichen RDBMSen, aber auch mit vielen NoSQL-Datenbanken nicht erreichbar.
- Schreiben wichtiger als Lesen! (Gemeint ist: Etwas in den Einkaufswagen tun.)
- Datenvolumen – nicht riesig.



Beispiel: Datenhaltung bei Amazon – Anforderungen (2)

- Deklarativer Zugriff
in voller Schönheit
anscheinend nicht so dringend.
- D. h. Verzicht auf Funktionalität
zugunsten von besseren
nichtfunktionalen Eigenschaften.



Die großen Themen der Vorlesung

- Themen dieser Vorlesung sind auch dann wichtig, wenn nicht explizit RDBMSe zum Einsatz kommen.
- Entwurf
 - subtile Unterschiede in den Bedeutungen der Konstrukte,
 - grundlegende Prinzipien bei der Transformation von Entwurfsergebnissen,
 - Was ist guter Entwurf?
- deklarativer Zugriff,
- Mehrbenutzerbetrieb,
- (einfache) Aussagen zur effizienten Arbeitsweise von Datenbanksystemen.

Einleitung

Welt
ohne DB

Datenbanken
Übersicht

Terminologie

Querying

DB –
Merkmale

Anwend.

Aufbau der Vorlesung

- Aufbau der Vorlesung (und des darüber hinausgehenden Lehrangebotes des Lehrstuhls) deshalb:
 - Erläuterung von Datenbank-Konzepten anhand von relationaler Datenbank-Technologie.
 - Praktischer Anteil in Form von Einlassungen zu SQL (konkreter Anfragesprache für RDBMSen) und entsprechender Übungen
 - Vorstellung ausgewählter NoSQL-Datenbanken.
 - Einblick in Forschungsthemen mit Bezug zur vorliegenden Thematik. (Anhand unserer eigenen Forschung. ;-))

Einleitung

Welt
ohne DB

Datenbanken
Übersicht

Terminologie

Querying

DB –
Merkmale

Anwend.

Weiterführende Vorlesungen, Auswahl

- Datenbankeinsatz
 - Alternativen zu relationalen Datenbanken
 - NoSQL-Datenbanken ausführlicher, insbesondere document-centric databases und Graph-Datenbanken.
- Analysetechniken für große Datenbestände
 - Techniken zum Finden von Auffälligkeiten und Besonderheiten in großen, komplex strukturierten Datenbeständen
 - Clustering, Association Rules, Klassifikation
 - für ‚herkömmliche‘ statische Daten, aber auch für hochdimensionale Daten und Datenströme
 - komplementäre Themen wie Data Warehousing oder Data Cleaning
- Datenhaltung in der Cloud
 - Tuning von Datenbanken, möglichst automatisiert.
 - Konsistenz für verteilte Datenbestände.

Einleitung

Welt
ohne DB

Datenbanken
Übersicht

Terminologie

Querying

DB –
Merkmale

Anwend.

Einbindung in Programme

- Nahtlose Einbindung von Querysprachen in Programme.
 1. Programmvariablen als Bestandteil der Anfrage.
 2. Queryergebnis ans Programm.

- Beispiel JDBC:

...

```
String query = "SELECT titel, preis, "  
+ "bestand FROM buch";
```

```
Statement stmt = con.createStatement ();  
ResultSet rs = stmt.executeQuery (query);
```

Einleitung

Welt
ohne DB

Datenbanken
Übersicht

Terminologie

Querying

DB –
Merkmale

Anwend.

„What about Performance?“

- Schein-Gegenargument:
„Generischer Code stets langsamer.“
- B-Baum Implementierung in MS-SQL
→ Weltrekord TPC-C.

Einleitung

Welt
ohne DB

Datenbanken
Übersicht

Terminologie

Querying

DB –
Merkmale

Anwend.

Datenbanken sind groß. (1)

- Warum dieser Aufwand?
Reichen einfache Lösungen wirklich nicht?
- Datenbanken sind groß, s. b. folgende Folien:
 - Größe einzelner Relationen,
 - Anzahl der Relationen
(mehrere Tausend bei SAP),
 - Anfragen sind komplex
(synthetisch generierte Anfragen greifen leicht
auf ein Dutzend oder mehr Relationen gleichzeitig zu).

Einleitung

Welt
ohne DB

Datenbanken
Übersicht

Terminologie

Querying

DB –
Merkmale

Anwend.

Zahlen – Bank

Große europäische Bank; Zahlen von 05/2005:

- > 70.000 Datenbanken
(zumeist DB2, Oracle, IMS, SQLServer)
- Größte Datenbank:
 - 25 Mio. Transaktionen pro Tag
(30 pro Sekunde),
 - Datenvolumen: 80 TB.
(1 TB – ca. 1 Milliarde Absätze mit Text)
- 99,7% Verfügbarkeit,
- 210 Mio. gedruckte Seiten pro Jahr,
- 60 Mitarbeiter nur für Datenbankadministration.

Einleitung

Welt
ohne DB

Datenbanken
Übersicht

Terminologie

Querying

DB –
Merkmale

Anwend.

Zahlen – SAP

- Anzahl Tabellen in SAP ERP: ca. 20.000,
(eher wenige große Tabellen; viele kleine, die den Charakter von .cfg oder .ini Files in einem Betriebssystem haben)
- Anzahl Nutzer
 - registrierte Benutzer: Tausende
 - “concurrent users”: 20-50 (i. Allg.)

Einleitung

Welt
ohne DB

Datenbanken
Übersicht

Terminologie

Querying

DB –
Merkmale

Anwend.

Datenbanken sind groß. (2)

- Unter diesen Umständen
 - spielt Reihenfolge der Zugriffe auf Relationen große Rolle (physische Datenunabhängigkeit),
 - sind logische Datenunabhängigkeit,
 - ausgefeilte Nebenläufigkeits-Zugriffsmechanismen,
 - Benutzerverwaltung etc. unumgänglich.

Einleitung

Welt

ohne DB

Datenbanken
Übersicht

Terminologie

Querying

DB –
Merkmale

Anwend.