

Datenbanken: Eine Einführung

Stefan Conrad & Leonie Selbach

Heinrich-Heine-Universität Düsseldorf

Institut für Informatik

Vorlesung im SS 2024

Organisatorisches und kurze Einführung

zwei aufeinander aufbauende Module

1. Modul „Datenbanken: Eine Einführung“

- 2 SWS VL, 2 SWS theoretische Übung
- 5 CP; mit Klausur
- ab PO 2021 Pflichtveranstaltung im 2. Fachsemester (BSc. Informatik), immer im Sommersemester

2. Modul „Datenbanken: Weiterführende Konzepte“

- setzt „Datenbanken: Eine Einführung“ voraus
- 2 SWS VL, 2 SWS praktische Übung
- 5 CP; benotetes Projekt („Hausarbeit“)
- ab dem 4. Fachsemester (BSc. Informatik), i.d.R. im Wintersemester

alle wichtigen Details auch in Ilias!

- Vorlesung:
 - Präsenzveranstaltung + *voraussichtlich* Aufzeichnung
 - Folien in Ilias: <https://ilias.hhu.de/>
- Übungen
 - Anmeldung für Gruppen: 12. - 22.4.
 - Start: 23./24.4.
- Fragen
 - allgemeine Fragen: Forum in Ilias oder Fragestunde
 - wenn es nur Sie betrifft: Ticketsystem (dbs@cs.uni-duesseldorf.de) oder Sprechstunde

- Tests
 - Jeden Montag 12 Uhr online
 - Bearbeitungszeit 1 Woche
 - Start: 15.4.
- Zulassungskriterien:
 - 50% der möglichen Punkte in Block 1 (Test 1 – 6)
 - 50% der möglichen Punkte in Block 2 (Test 7 – 12)
 - Altzulassungen bleiben bestehen
- Abschlussprüfung (1. Klausur am 05.08.2024)

	Tag	Zeit	Raum	
Vorlesung	Donnerstag	12:30 – 14:00	HS 6J	S. Conrad & L. Selbach
Übungen	Dienstag	8:30 – 10:00		L. Selbach & Tutor:innen
	Dienstag	14:30 – 16:00		
	Mittwoch	14:30 – 16:00		
Fragestunde	Montag	14:30 – 16:00	webex	L. Selbach
Sprechstunde	Mittwoch	10:00 – 11:30	webex	L. Selbach

Modul „Datenbanken: Eine Einführung“

1. Grundlegende Konzepte und Architekturen
2. Datenbankmodelle für den Entwurf (\rightarrow ER)
3. Datenbankmodelle für die Realisierung
4. Datenbankentwurf
5. Grundlagen von Anfragen
6. Relationale Datenbanksprachen
7. Relationaler Datenbankentwurf
8. Transaktionsverwaltung

Ausblick: Modul „Datenbanken: Weiterführende Konzepte“

1. Datendefinition (SQL DDL)
2. Datenbank-Anwendungsprogrammierung
3. Web & Datenbanken
4. Sichten, Datenschutz
5. Integrität und Trigger
6. Dateiorganisation und Indexstrukturen
7. DB-Basisoperationen (und ihre Realisierung)
8. Anfrageoptimierung

Danksagung:

Diese Folien basieren im Wesentlichen auf dem Buch von Heuer, Saake & Sattler (s.u.). Die Buchautoren haben dazu einen umfangreichen Foliensatz entwickelt.

Im Laufe der letzten Jahre wurde der Foliensatz immer wieder etwas überarbeitet.

Allen, die an der Erstellung und Bearbeitung der Folien mitgewirkt haben, sei an dieser Stelle für ihre Mitwirkung gedankt!

Hinweis: Folien sind eine Präsentationshilfe und können daher keine vollständigen Erklärungen enthalten. Zum Lernen sind i.d.R. eigene Notizen zu den Folien und/oder das Hinzuziehen von Lehrbüchern erforderlich.

Literatur (Auswahl)

- Heuer, A., Saake, G., Sattler, K.-U.; *Datenbanken — Konzepte und Sprachen*. 6. Aufl., mitp-Verlag, Bonn, 2018
- Kemper, A., Eickler, A.; *Datenbanksysteme — Eine Einführung*. 10. Aufl., DeGruyter, 2015
- Vossen, G.; *Datenbankmodelle, Datenbanksprachen und Datenbankmanagement-Systeme*. 5. Aufl., DeGruyter, München, 2008
- Kudraß, T. (Hrsg.); *Taschenbuch Datenbanken*. Carl Hanser Verlag, 2. Aufl., 2015.
- Faeskorn-Woyke, H.; Bertelsmeier, B.; Riemer, P.; Bauer, E.; *Datenbanksysteme: Theorie und Praxis mit SQL2003, Oracle und MySQL*. Pearson Studium, 2007.
- Elmasri, R.; Navathe, S.B.; *Fundamentals of Database Systems*. 7th Edition, Addison-Wesley / Pearson, 2016
- Date, C.J.; *An Introduction to Database Systems*. 8. Aufl., Addison-Wesley / Pearson, 2004
- Ullman, J.D.; Garcia-Molina, H.; Widom, J.; *Database Systems – The Complete Book*. International Edition, 2. Aufl., Pearson, 2013

- Querbezüge zu anderen Bereichen der Informatik:
 - Modellierung, Datenstrukturen, Theorie, Betriebssysteme/Verteilte Systeme, Sicherheit, Multimedia, ...
- Datenbanksysteme als Basis moderner Softwaresysteme:
 - Web-basierte Systeme (eBay, Amazon, Facebook, ...)
 - ERP-Systeme (SAP R/3), CRM-Systeme, Finanzsysteme
 - administrative Anwendungen
 - wiss. Anwendungen (NASA Earth Observation System, Human Genom Project, ...)

- DB-Kenntnisse unverzichtbar für Informatik-Berufe:
 - Administration, Planung/Entwurf, Entwicklung, Nutzung
- große Herausforderungen:
 - Verwaltung von Daten im TByte/PByte-Bereich, viele Nutzer*innen
 - weltweit verteilte Datenbestände
 - Multimedia-Inhalte
 - Hochverfügbarkeit, Sicherheit
 - Effizienz / Performanz
 - Anpassung an neue Hardware

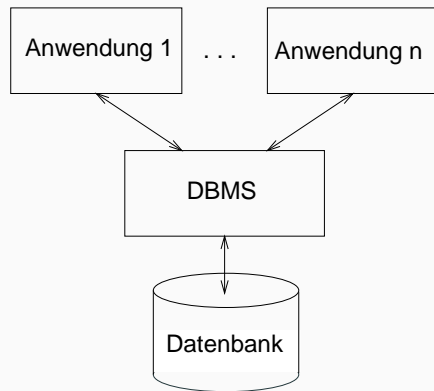
Ohne Datenbanken: Datenredundanz

- Basis- oder Anwendungssoftware verwaltet ihre eigenen Daten in ihren eigenen (Datei-)Formaten
 - Textverarbeitung: Texte, Artikel und Adressen
 - Buchhaltung: Artikel, Adressen
 - Lagerverwaltung: Artikel, Aufträge
 - Auftragsverwaltung: Aufträge, Artikel, Adressen
 - CAD-System: Artikel, Technische Bausteine
- Daten sind **redundant**: mehrfach gespeichert;
Probleme:
 - Verschwendung von Speicherplatz
 - „Vergessen“ von Änderungen
 - keine zentrale, „genormte“ Datenhaltung

- Andere Software-Systeme können große Mengen von Daten nicht effizient verarbeiten.
- Mehrere Benutzer:innen oder Anwendungen können nicht parallel auf den gleichen Daten arbeiten, ohne sich zu stören.
- Anwendungsprogrammierer:innen / Benutzer:innen können Anwendungen nicht programmieren / benutzen, ohne
 - interne Darstellung der Daten
 - Speichermedien oder Rechnerzu kennen (**Datenunabhängigkeit** nicht gewährleistet).
- Datenschutz (kein unbefugter Zugriff) und Datensicherheit (kein ungewollter Datenverlust) sind nicht gewährleistet

- Anfang 60er Jahre: elementare Dateien, anwendungsspezifische Datenorganisation (geräteabhängig, redundant, inkonsistent)
- Ende 60er Jahre: Dateiverwaltungssysteme (SAM, ISAM) mit Dienstprogrammen (Sortieren) (geräteunabhängig, aber redundant und inkonsistent)
- 70er Jahre: Datenbanksysteme (Geräte- und Datenunabhängigkeit, redundanzfrei, konsistent)

- **DBS:** Datenbanksystem
DBMS + Datenbank
- **DBMS:** Datenbank-Management-System
Software zur Verwaltung der Datenbank



- Grundmerkmale DBMS
 - verwalten persistente (langfristig zu haltende) Daten
 - verwalten große Datenmengen effizient
 - Datenbankmodell, mit dessen Konzepten alle Daten einheitlich beschrieben werden (Integration)
 - Operationen und Sprachen (DDL, DML, ...) deskriptiv, getrennt von einer Programmiersprache
 - Transaktionskonzept, Concurrency Control: logisch zusammenhängende Operationen atomar (unteilbar), Auswirkungen langlebig, können parallel durchgeführt werden
 - Datenschutz, Datenintegrität (Konsistenz), Datensicherheit

- 1970: Ted Codd (IBM) → Relationenmodell als konzeptionelle Grundlage relationaler DBS
- 1974: System R (IBM) → erster Prototyp eines RDBMS
 - zwei Module: RDS, RSS; ca. 80.000 LOC (PL/1, PL/S, Assembler), ca. 1,2 MB Codegröße
 - Anfragesprache SEQUEL
 - erste Installation 1977
- 1975: University of California at Berkeley (UCB) → Ingres
 - Anfragesprache QUEL
 - Vorgänger von Postgres, Sybase, ...
- 1979: Oracle Version 2
 - erstes kommerziell verfügbares RDBMS