

Einführung in Datenbanken

Kapitel 0: Organisatorisches

Prof. Dr. Stefan Brass

Martin-Luther-Universität Halle-Wittenberg

Wintersemester 2023/24

<http://www.informatik.uni-halle.de/~brass/db23/>

Inhalt

- 1 Inhalte
- 2 Motivation
- 3 Organisatorisches
- 4 Literatur
- 5 Software
- 6 Bitten

Ist dies die richtige Vorlesung? (1)

- Die Vorlesung „Datenbanken I“ (10 LP) wurde aufgespalten:
 - Einführung in Datenbanken (5 LP) (Wintersemester)
 - Datenbank-Programmierung (5 LP) (Sommersemester)

- Wenn Ihre Prüfungsordnung „Datenbanken I“ fordert, werden beide Module zusammen dafür angerechnet.

Bei besonderen Schwierigkeiten durch diese Aufspaltung kontaktieren Sie mich.

- Für Informatiker und Bioinformatiker ist nach der neuen Prüfungsordnung nur noch diese Vorlesung Pflicht.

Der zweite Teil, „Datenbank-Programmierung“, ist Wahlpflicht. Wenn Sie sich im Master in Datenbanken spezialisieren wollen, sollten Sie das hören.

- Für Wirtschaftsinformatiker ist der erste Teil Pflicht, der zweite fast (Wahlbereichspflichtbereich 1.3: 4 aus 5).

Ist dies die richtige Vorlesung? (2)

- Dies ist nicht die Vorlesung
„Einführungen in Datenbanken und das WWW (ASQ)“,
die u.a. auch im Ergänzungsstudiengang Medienbildung
belegt werden kann.

Die ASQ-Vorlesung wird momentan leider nicht angeboten.
Falls Sie diese (anspruchsvollere) Vorlesung als Ersatz belegen möchten,
sprechen Sie bitte mit dem Dozenten.
- Falls Sie Lehramt an Sekundarschulen mit Erweiterungsfach
Informatik studieren, benötigen Sie eine mündliche Prüfung.

Eine Prüfung über beide Teile statt der zwei Klausuren. Auch in diesem
Fall bitte den Dozenten ansprechen (gegen Ende).

Themen

Diese Vorlesung ist eine grundlegende Einführung in die Anwendung relationaler Datenbank-Managementsysteme.

Inhalte:

- Grundlegende Begriffe, relationales Datenmodell
- Etwas Logik (mit Anwendungen für Datenbanken/SQL)
 - Man kann viele Datenmodelle und Datenbanksprachen mit den Begriffen der Logik verstehen. Logik hilft, SQL besser zu nutzen.
- Die Datenbanksprache SQL (Schwerpunkt der Vorlesung)
- Relationale Algebra (kurz)
- Einführung in DB-Entwurf (ER-Modell, BCNF, → DB IIA)
- Updates, Transaktionen (sehr kurz, → DB-Programmierung)

Fortsetzung: Datenbank-Programmierung

Geplante Inhalte für das Modul im Sommer:

- Fortgeschrittenes/Modernes SQL
 - Insbesondere für Data Warehouse Anwendungen.
 - Auch rekursive Anfragen in SQL (ggf. mit Ausblick auf Datalog)
- Transaktionen (Recovery, Mehrbenutzerbetrieb)
 - Synchronization paralleler Zugriffe auf die Datenbank
- Datenbank-Schnittstellen aus Programmiersprachen, JDBC
- Integritäts-Überwachung, Serverseitige Programmierung, Trigger
- Zugriffsrechte, Datenschutz, Sichten
- Projekt mit Git, Maven, ggf. Docker (Prüfung: Hausarbeit)

Inhalt

- 1 Inhalte
- 2 Motivation**
- 3 Organisatorisches
- 4 Literatur
- 5 Software
- 6 Bitten

Motivation (1)

Warum sind Datenbanken wichtig?

- „Informations-Gesellschaft“. Information ist eines der wichtigsten Aktiva vieler Wirtschaftsunternehmen.
- Datenbanken sind das Speichermedium für strukturierte Informationen (z.B. Tabellen).
- Sie werden im Berufsleben in der einen oder anderen Form mit Datenbanken zu tun haben.
- Viele Webseiten beziehen die Inhalte aus Datenbanken. Jeder Webshop hat als Grundlage eine Datenbank.
- „Big Data“ ist in aller Munde.
Häufig werden die Daten für Analysen in Datenbanken abgelegt (teils „NoSQL“).
- Verständnis der Welt durch Strukturierung von Information.

Motivation (2)

Warum machen Datenbanken Spaß?

- Interessante Verbindung von Theorie und Praxis, z.B. beruht SQL auf der Prädikatenlogik 1. Stufe.

Viele DB-Begriffe sind so einfach, dass noch formal zu definieren.

- Zumeist sehr anschaulich, anwendungsorientiert.
- Relativ wenig komplexe Programmierung, wenn man nicht gerade ein neues DBMS implementiert.
- Interessante Algorithmen und Datenstrukturen, gute Anwendung für Parallelität.

Performance ist sehr wichtig, wenn man große Datenmengen verwalten muss (z.B. Hunderte von Terabytes), und der Server viele (z.B. > 10 000) Transaktionen pro Sekunde durchführen soll.

- Noch nicht von Microsoft dominiert.

Motivation (3)

Aufgabe:

- Wo begegnen Ihnen im normalen Leben Datenbanken?
Welche enthalten Informationen über Sie?

- Supermarkt (Scannerkasse)

- _____
- _____
- _____
- _____
- _____

Inhalt

- 1 Inhalte
- 2 Motivation
- 3 Organisatorisches**
- 4 Literatur
- 5 Software
- 6 Bitten

Ansprechpartner (1)

Vorlesung und Übungsgruppe 2 (Di 10-12): Prof. Dr. Stefan Brass

- Email: brass@informatik.uni-halle.de

Betreff-Zeile sollte mit [db23] beginnen, möglichst aussagefähig.

- Büro: Von-Seckendorff-Platz 1, Raum 313

Ich mache viel „Home Office“, bin dort also eher selten zufällig anzutreffen.

- Sprechstunde (Büro, 0345/55-24740): Montags, 12⁰⁰–13⁰⁰

Bitte nur mit Maske kommen. Ich bin zu dieser Zeit auch telefonisch zu erreichen. Bei Bedarf Videokonferenz (per EMail absprechen).

- Frühere Unis: Braunschweig, Dortmund, Hannover, Hildesheim, Pittsburgh, Gießen, Clausthal.

- Oracle8 Certified Database Administrator (aktuell: 19C,21C)
IBM Certified Advanced DBA (DB2 UDB 8.1) (aktuell: 12)

Ansprechpartner (2)

Übungsgruppe 1 (Di 14–16): PD Dr. Alexander Hinneburg

- Büro: Von-Seckendorff-Platz 1, Raum 314
- Telefon: 0345/55-24732
- Email: hinnebur@informatik.uni-halle.de

Sekretärin: Ramona Vahrenhold

- Büro: Von-Seckendorff-Platz 1, Raum 324
- Telefon: 0345/55-24750, Fax: 0345/55-27333
- Email: vahrenho@informatik.uni-halle.de

Zeit und Ort (1)

Vorlesung (2 SWS):

- Montags, 10¹⁵–11⁴⁵, Raum 3.07.
- Es gibt (meistens) eine Videoaufzeichnung, aber ist empfohlen, an der Vorlesung in Präsenz teilzunehmen.

StudIP-Eintrag der Vorlesung, Reiter „Opencast“. Es gibt etwas zeitliche Verzögerung, erfahrungsgemäß steht die Aufzeichnung ab ca. 17 Uhr zur Verfügung. Etwas später (ggf. am nächsten Tag) auch „geschnitten“. Bei technischen Problemen ist nicht garantiert, dass die Aufzeichnung wiederholt wird. Falls Sie die Vorlesung verpasst haben, sollten Sie die Aufzeichnung zeitnah anschauen, und es nicht zu lange verschieben.

- Fragen sind willkommen!

Sie sind auf der Aufzeichnung praktisch nicht zu hören. Ich sollte deshalb daran denken, die Frage nochmals zu wiederholen.

Zeit und Ort (2)

Übung (2 SWS):

- Zwei Gruppen (Anmeldung: StudIP):

Nr	Tag	Zeit	Raum	Dozent
1	Dienstag	14 ¹⁵ –15 ⁴⁵	3.34 (Pool)+online	Hinneburg
2	Dienstag	10 ¹⁵ –11 ⁴⁵	3.04 (Hörsaal)	Brass

- Die Übungen beginnen morgen, 10.10.2023.

Es werden auch organisatorische Dinge ausführlich besprochen.

- Wenn Sie an der Präsenzübung von Gruppe 1 im PC-Pool (3.34) teilnehmen wollen, sollten Sie sich im Wiki des StudIP-Eintrags anmelden.

Es gibt nur 22 Plätze und momentan 37 Teilnehmer in StudIP.

Zeit und Ort (3)

Übung, Forts.:

- Zur Teilnahme an der Online-Übung von Gruppe 1 gehen Sie auf den StudIP-Eintrag der Übungsgruppe und wählen dort den Reiter „Meetings“. Es wird **BigBlueButton** verwendet.

Eigene Installation der Universität: [<https://mluconf.uni-halle.de>].

- Bei Gruppe 2 werden eventuell Teile aufgezeichnet (natürlich ohne das Vorrechnen der Übungsaufgaben durch Studierende, es fehlen also ganz wichtige Teile!).

Das ist nur für Notfälle gedacht, wenn Sie eine Übung verpasst haben.

Schauen Sie ggf. in den StudIP-Eintrag von Gruppe 2, Reiter „Opencast“.

- Es ist besser, an Übungen „life“ teilzunehmen.

Dann können Sie auch Fragen stellen. Dafür ist die Übung ja da.

Erwerb der Leistungspunkte

Studienleistung:

- Mindestens 50% der Hausaufgabenpunkte.

Die Hausaufgaben sind einzeln zu bearbeiten.

Es gibt voraussichtlich auch Präsenzaufgaben in den Übungen, diese geben auch Hausaufgabenpunkte und werden mitgerechnet. Die Präsenzaufgaben können in Gruppen bearbeitet werden.

- Aktive Mitarbeit in den Übungen (u.a. Vorrechnen von HAen)

Prüfung:

- Voraussichtlich elektronische Klausur in Präsenz am 26.02.2024 (??) und am 08.07.2024.

Bitte achten Sie auf mögliche Änderungen und melden Sie Terminkonflikte frühzeitig. Beim erstem Termin sind 5 Blätter/10 Seiten Notizen erlaubt, beim zweitem Termin 3 Blätter/6 Seiten (weil mehr Vorbereitungszeit).

Plagiage, Datenschutz-Hinweis

- Die Hausaufgaben sind einzeln zu bearbeiten.

Selbstverständlich sollen Sie einander helfen, aber nicht eine Lösung zum abschreiben geben.

- Bei Plagiaten gibt es 0 Punkte für alle Beteiligten (d.h. auch Autor/Autorin des Originals).

- Im Wiederholungsfall droht Schlimmeres.

Z.B. Ausschluss von der Übung/Verweigerung der Studienleistung.

- **Datenschutz-Hinweis:** Wir loggen die Zugriffe auf den Adminer (Web-Schnittstelle zum Ausprobieren der Anfragen) und verwenden die Daten zur Plagiatskontrolle und für ein Forschungsprojekt.

ChatGPT

- ChatGPT (KI-Chatbot) hat die Situation verändert.
- Sie könnten sich von diesem Programm SQL-Anfragen schreiben lassen (**Wir raten ab!**).
- Wir können das vermutlich nur selten zweifelsfrei nachweisen.

Manchmal schon, wenn ChatGPT ganz merkwürdige Lösungen produziert. Außerdem haben Sie das Risiko, untereinander des Plagiats beschuldigt zu werden, wenn mehrere Studierende ChatGPT nutzen. ChatGPT produziert aber unterschiedliche Lösungen, auch für die gleiche Eingabe.
- Manche der generierten Anfragen sind falsch.

Aber für 50% der Übungspunkte würde es sicher reichen.
- **Es bleibt also eine ganz wesentliche Kompetenz, SQL-Anfragen auch selbst schreiben zu können.**

Sie müssen ja wenigstens die Korrektheit kontrollieren.

Vorschlag/Idee/Forschungsprojekt (1)

- Sie können bei den Hausaufgaben (nicht bei der Klausur!) beliebig abschreiben und ChatGPT verwenden.
- Bedingung ist aber, dass Sie die Quelle angeben (also auch den Namen des/der Mitstudierenden, der/die die originale Lösung verfasst hat).

Oder eben ChatGPT oder eine andere KI, oder eine Webseite, oder eine Sammlung von Lösungen aus einem früheren Semester. Wir werden die Daten natürlich nur anonymisiert für unser Forschungsprojekt verwenden. Sie können auch „(modifiziert)“ (bei stilistischen Verbesserungen) oder „(überarbeitet)“ (bei inhaltlichen Veränderungen) dazu schreiben. Mehr Details morgen in der Übung und auf der Webseite zur Vorlesung.

- Beispiel:

-- Quelle: Müller, Monika (modifiziert)

Vorschlag/Idee/Forschungsprojekt (2)

- Wenn Sie keine Quelle angeben, verstehen wir das als Selbständigkeitserklärung.
- Dann gibt es eine normale Plagiatskontrolle, auch ein Vergleich mit Lösungen von ChatGPT.
- Mit der Angabe des Autors/der Autorin schützen Sie auch Ihren Mitstudenten/Ihre Mitstudentin:
 - Er/Sie würde, falls die Kopie erkannt wird, und die Quellenangabe fehlt, auch 0 Punkte bekommen.
 - Dann wäre es ein Täuschungsversuch.
- **Wenn Sie zu viele Anfragen nicht selbst entwickeln, werden Sie bei der Klausur durchfallen!**

Aber die Studienleistung bekommen Sie. Wir ziehen keine Punkte für die Quellenangabe ab.

Inhalt

- 1 Inhalte
- 2 Motivation
- 3 Organisatorisches
- 4 Literatur**
- 5 Software
- 6 Bitten

Vorlesungs-Materialien

- Das alte Skript zu „Datenbanken I“ hatte ca. 1100 Folien.

Es ist verfügbar unter folgender Adresse:

<http://www.informatik.uni-halle.de/~brass/db17/>

- Im letzten Jahr gab es 1022 Folien (fast ein Skript).

Ein Teil sind nicht prüfungsrelevante Anmerkungen zu speziellen DBMS.

Ein Teil sind aber auch Hinweise zu Fehlern und Missverständnissen, die mir aufgefallen sind. Es gibt jedes Jahr eine Überarbeitung.

- Die Folien werden kurz vor der Vorlesung auf folgende Webseite gestellt:

<http://www.informatik.uni-halle.de/~brass/db23/>

- Webseite enthält auch: Alte Klausuren, Links.
- Forum in StudIP (z.B. für Fragen zu Hausaufgaben).

Bücher sind wichtig!

- Wir sind an einer Uni.

Ziel ist auch die kritische Auseinandersetzung mit dem Stoff (inklusive Stoffauswahl). Eigenes Denken und Verstehen statt Auswendiglernen!

Natürlich muss man am Ende syntaktische Details von SQL können.

Aber Datenbanken sind viel mehr als nur SQL, und auch da kann man sich ja fragen, ob es nicht bessere Anfragesprachen geben könnte.

- Man braucht mehrere Quellen.

Nur aus unterschiedlichen Blickwinkeln sieht man dreidimensional.

- Ich empfehle dringend, ein Lehrbuch zu beschaffen, und in Teilen zu lesen (reicht oft auch für Master-Vorlesungen).

Die Modulbeschreibung enthält hierfür 45–60 Stunden Selbststudium.

Folien sind kein Lehrbuch, und unterschiedliche Lernformen ergänzen sich.

Je nach Vorwissen/Interessen sind unterschiedliche Bücher ansprechend.

- Es sind auch Online-Quellen auf der Webseite verlinkt.

Lehrbücher (1)

- Alfons Kemper / André Eickler:
Datenbanksysteme. Eine Einführung. 10. Auflage.
De Gruyter, 2015, ISBN 3110443759, 880 Seiten, 49.95 €.
- Gunter Saake, Kai-Uwe Sattler, Andreas Heuer:
Datenbanken. Konzepte und Sprachen. 6. Aufl.
MITP Bonn, 2018, ISBN 978-3958457768, 800 Seiten, 39.99 €.
- Gottfried Vossen:
Datenbankmodelle, Datenbanksprachen und
Datenbankmanagementsysteme. 5. Aufl.
Oldenbourg, 2008, 3-486-27574-7, 842 Seiten, 64.95 €.

Lehrbücher (2)

- Ramez Elmasri / Shamkant B. Navathe:
Fundamentals of Database Systems, 7th Ed.

Pearson India, 2017, ISBN 933258270X, ca. 29.00 €

Global Ed.: Prentice Hall, 2016, ISBN 1-292-09761-2, 1272 S., 39.16 €

Deutsch: Bachelorausgabe (Teile fehlen/nur online — Übersetzung fragwürdig):

Pearson, 3. Aufl., 2009, ISBN 386894012X, 560 Seiten, 34.95 €.

- Hector Garcia-Molina, Jeffrey D. Ullman, Jennifer Widom:
Database Systems: The Complete Book, 2nd Ed.

Prentice Hall, 2008/13, ISBN 978-1292024479, 1203/1140 pages, 61.20 €.

- Raghu Ramakrishnan, Johannes Gehrke:
Database Management Systems, 3rd Ed.

McGraw Hill, 2002, ISBN 0-07-123151-X, 1065 Seiten, 65.99 €.

Lehrbücher (3)

- Philip M. Lewis, Arthur Bernstein, Michael Kifer:
Database and Transaction Processing, 2nd Ed.

Addison-Wesley, 2005, ISBN 0-321-26845-8, 1235 pp., 148.99 €.

TB: ISBN 0-321-31256-2, 76.99 €. Es gibt „Introductory Version“.

- Jim Melton, Alan R. Simon: SQL: 1999 —
Understanding Relational Language Components.

Morgan Kaufmann, 2. Aufl., 2001, ISBN 1558604561, 928 pp, 63.23 €.

- Georg Lausen:
Datenbanken. Grundlagen und XML-Technologien.

Spektrum/Elsevier, 2005, ISBN 3-8274-1488-1, 286 Seiten, nur noch
gebraucht.

Inhalt

- 1 Inhalte
- 2 Motivation
- 3 Organisatorisches
- 4 Literatur
- 5 Software**
- 6 Bitten

PostgreSQL: (Objekt-)Relationales DBMS

- Verbreitetes Open Source Datenbank-Managementsystem (DBMS) mit umfangreicher SQL-Unterstützung.

Postgres war Vorreiter der objektrelationalen Technologie.

Die Hausaufgaben sollten allerdings möglichst in Standard-SQL gelöst werden.

Bei der Klausur kann man voraussichtlich Anfragen mit PostgreSQL testen.

- [<https://www.postgresql.org/>]

Läuft unter Linux, Windows, MacOS, BSD Unix, Solaris.

- Wir bieten eine Adminer Web-Schnittstelle zum Üben von SQL ohne eigene Installation. [<https://www.adminer.org/>]

[[https://dbs.informatik.uni-halle.de/edb?](https://dbs.informatik.uni-halle.de/edb?pgsql=db&username=&db=postgres&ns=)

[pgsql=db&username=&db=postgres&ns=](https://dbs.informatik.uni-halle.de/edb?pgsql=db&username=&db=postgres&ns=)]

Die Login-Daten stehen im [StudIP](#)-Eintrag der Vorlesung, Reiter „Informationen“.

- Spätestens in der Folge-Vorlesung „DB-Programmierung“ braucht man eine eigene Installation (ggf. mit Docker).

Weitere Software

- Es gibt noch viele weitere Datenbank-Managementsysteme, z.B. MySQL, Oracle, Microsoft SQL Server.

Im Internet gibt es auch Webseiten zum Ausprobieren dieser Systeme.

Auch von kommerziellen Systemen wie Oracle und Microsoft SQL Server gibt es eingeschränkte „Express Editions“, die kostenlos sind.

- In der ersten Übung und im Kapitel 2 wird das noch weiter diskutiert, auch die Bedeutung portabler Programmierung.
- Zum Ausprobieren der relationalen Algebra verwenden wir RelaX ([Artikel auf BTW 2019](http://dbis-uibk.github.io/relax/calc/gist/8dc2652578ee12ae756a234c4cf21b3f)).
[\[http://dbis-uibk.github.io/relax/calc/gist/8dc2652578ee12ae756a234c4cf21b3f\]](http://dbis-uibk.github.io/relax/calc/gist/8dc2652578ee12ae756a234c4cf21b3f)
- Für den DB-Entwurf mit Entity-Relationship-Diagrammen wird der „Oracle SQL Developer Data Modeler“ verwendet.
[\[https://www.oracle.com/de/database/technologies/appdev/datamodeler.html\]](https://www.oracle.com/de/database/technologies/appdev/datamodeler.html)

Inhalt

- 1 Inhalte
- 2 Motivation
- 3 Organisatorisches
- 4 Literatur
- 5 Software
- 6 Bitten**

Aktive Mitarbeit, Verbesserung der Lehre

- Fragen sind sehr willkommen!

In der Präsenz-Version der Vorlesung und vor allem den Übungen.

Eventuell im Forum in StudIP, nur selten per EMail (ich bekomme sehr viele EMails, und schreibe teils lange Antworten, was mir viel Zeit kostet).

- Vorschläge zur Verbesserung der Lehrveranstaltung sind willkommen.

Das gern per EMail. Z.B. Hinweise auf Tippfehler, veraltete Angaben, unklare Stellen, u.s.w. Auch Hinweise auf nützliche Webseiten.

- Beteiligen Sie sich in den Übungen mit Wortbeiträgen.

Im Notfall muss ich irgendwen aufrufen, der sich nicht gemeldet hat, und die ganze Sache wird zäh und für alle (auch mich) wenig erfreulich.

- Geben Sie bei der Videokonferenz Ihren echten Namen an.

Bedeutung praktischer Übung

- Ich möchte, dass Sie am Ende dieser Vorlesung SQL problemlos beherrschen, und zwar nicht nur simple Anfragen, sondern auch komplexere.

Z.B. mit NOT EXISTS, GROUP BY, UNION, CASE, LEFT JOIN.

- Neben theoretischem Verständnis braucht man dazu auch praktische Erfahrung mit einem konkreten DBMS.

Wie bei jeder Programmiersprache. Zum Teil ist es Handwerk. Man muss üben.

- Klären Sie Fehler vollständig auf. Fehler sind nützlich!

Kopieren Sie sich Anfrage und Fehlermeldung bzw. Ergebnis und fragen Sie in der nächsten Übung, wenn Sie es nicht selbst aufklären können.

Wenn jemand in der Klausur mit einer simplen Fehlermeldung von PostgreSQL nicht umgehen kann, die er eigentlich schon viele Male gesehen haben muss, frage ich mich doch, ob er die Hausaufgaben selbst gemacht hat.