Datenbanken IIB: Implementierung von DB-Managementsystemen

(Wintersemester 2013/14)

Prof. Dr. Stefan Brass
Institut für Informatik

Übungsleiter: Prof. Dr. Stefan Brass, Dipl.-Inform. Heike Stephan

Themen

Wie funktioniert ein DBMS? ($\rightarrow$ Performance Tuning)

- Platten, RAID-Systeme, SANs, Pufferung/Caching
- Speicherverwaltung, Physische Parameter bei der Tabellendeklaration in Oracle
- Indexe: B-Bäume, ggf. Cluster, Hash, Bitmap
- Anfrage-Auswertungspläne, Algorithmen für Operationen der relationalen Algebra
- Anfrage-Optimierung
- Ggf. Backup und Recovery

Motivation

Warum diese Vorlesung hören?

- Neugierde: Wie funktionieren DBMS eigentlich?
- Damit auch besseres Verständnis für Datenbanksysteme, effektivere Anwendung.
- Teil der Vorbereitung für
  - ◇ Beruf "Datenbankadministrator" (DBA).
  - ◇ Zertifizierung als Oracle DBA.
  - ◇ Beruf "Berater für Datenbank-Tuning".

    Echte Experten sollen sehr gut bezahlt werden, weil sie den Firmen das Geld für eine größere Hardware sparen.

Voraussetzungen

Dies ist eine fortgeschrittene DB-Vorlesung. Vorausgesetzt werden:

- Relationales Modell
- SQL (komplexere Anfragen, `CREATE TABLE`, `GRANT`)
- Operationen der relationalen Algebra
- Programmierkenntnisse, Datenstrukturen
- ER-Diagramme (Grundkenntnisse)
- (mehr oder weniger:) Englische Sprachkenntnisse

Nicht vorausgesetzt: Datenbanken IIA

Zeit und Ort

Vorlesung:

- Donnerstags, 10^{15}–11^{45}, Raum 3.04

Übung (beginnt 10.10.2013):

- Nur eine Gruppe: Donnerstags, 12^{15}–13^{45}
- Tafelübung: Raum 3.04 (ca. 45 min)
- Praktische Übung: PC-Pool, Raum 3.32

  Die Aufteilung zwischen den beiden Übungsformen ist variabel.
- 20 Plätze (bitte in Punkte-DB anmelden)

  Und auch in StudIP mindestens für die Vorlesung eintragen.

Zeitliche Belastung

- Diese Vorlesung hat 5 Leistungspunkte.

  Auch "credit points" genannt.
- Entspricht 150 Stunden studentischer Arbeitszeit:

Lernform	SWS	Stunden
Vorlesung	2	30
Selbststudium	0	60/45
Tafelübung	1	15
Praktische Übung (z.T. HA)	1	15
Hausaufgaben	0	30/45

Ansprechpartner (1)

Dozent: Prof. Dr. Stefan Brass

- Büro: Von-Seckendorff-Platz 1, Raum 313
- Telefon: 0345/55-24740
- Email: `brass@informatik.uni-halle.de`
- Sprechstunde: Montags, 12^{00}–13^{00}
- Frühere Unis: Braunschweig, Dortmund, Hannover, Hildesheim, Pittsburgh, Gießen, Clausthal.
- Oracle8 Certified Database Administrator
  IBM Certified Advanced DBA (DB2 UDB 8.1).

Ansprechpartner (2)

Übungsleiterin (ab ca. 01.11.): Heike Stephan

- Büro: Von-Seckendorff-Platz 1, Raum 315
- Telefon: 0345/55-24776
- Email: `heike.stephan@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`

Webseiten

`http://www.informatik.uni-halle.de/~brass/dbi13w/`

- Aktuelle Ankündigungen
- Folien der Vorlesung (in Englisch)

  Die Folien werden vor der entsprechenden Vorlesung (nach Fertigstellung bzw. Überarbeitung) ins Netz gestellt. Kleinere Korrekturen oder Ergänzungen sind auch noch später möglich.
- Alte Klausuren
- Verweise auf Literatur im WWW
- Punkte-Datenbank (s.u.)
- Hausaufgaben, Informationen zur Übung

Prüfung (1)

Modulvorleistung/Studienleistung:

- Bedingung für erfolgreichen Abschluss des Moduls.

  Auch für Diplom-Studierende (für Ausnahme Dozent fragen).
- Mindestens ca. 2/3 der Hausaufgabenblätter bearbeitet und abgegeben (per EMail, Betreff `[dbi13w]`).

  Hausaufgaben werden nicht korrigiert, aber in der Übung besprochen. 2er-Gruppen sind erlaubt, jeder muß vorrechnen können. Zu ähnliche oder offensichtlich schlechte Lösungen zählen als nicht bearbeitet.
- Regelmäßige und aktive Mitarbeit in den Übungen.

  Höchstens drei Mal fehlen oder mit Übungsleiter besprechen. Hausaufgaben in der Übung an der Tafel präsentieren, dabei auch Fragen zum Umfeld beantworten. Präsenzaufgaben bearbeiten. Mitdiskutieren.

Prüfung (2)

Idee für Teil der Hausaufgaben:

- Wir könnten ein Stück eines DBMS in C++ programmieren.

  Z.B. Pufferverwaltung, Datenstrukturen für Relationen und Tupel (Satzschnittstelle). Bis zu SQL würden wir nicht kommen, äußerstenfalls eine relationale Algebra (Auswertungspläne). Parallele Zugriffe und Transaktionsverwaltung (Rollback, Backup&Recovery) würden wir auch weglassen. Eventuell könnte man die Aufgaben auch aufteilen, wenn man in der Übung gemeinsam die Schnittstelle entwirft.
- Alternativ: Stück eines Tripelspeichers (RDF)?
- Außerdem sind Experimente und die Durchführung typischer DBA-Aufgaben mit Oracle nötig.

Prüfung (3)

Modulleistung:

- Klausur: 27.03.2014, 10^{00}–12^{00} (Vorschlag).

  Der Termin kann sich eventuell noch ändern. Bitte achten Sie auf Ankündigungen auf der Webseite. Eventuell mündliche Prüfung.
- Nachholtermin: mündliche Prüfung, ca. Mai. 2014
- Bei der Klausur sind drei DIN A4-Blätter mit Notizen erlaubt, keine Bücher, Aktenordner, Rechner.
- Anwendung, Verstehen, kein Auswendiglernen.

  Z.B.: Speicherplatz für Tabelle bestimmen, günstigen Index für gegebene Anfrage auswählen, Auswertungsplan für SQL-Anfrage angeben, Zugriffszeiten schätzen. Kurzer Aufsatz zu Frage.

Wichtige Hinweise

- Sie müssen sich zum Modul (bis 23.10.2013) und zur Klausur anmelden.

  Falls Prüfung (Informatik/Bioinformatik Master, Wirtschaftsinformatik Diplom): Prüfungsamt oder Löwenportal. Falls nur Übungsschein: Beim Übungsleiter. Ein Schein ist keine Voraussetzung für eine mündliche Diplomprüfung, aber eine nützliche Übung.

- Für Master-Studierende der Wirtschaftsinformatik: Das Modul kann nur nach Beratung durch Prof. Dr. Mellouli belegt werden.
- Falls Ihre Zensur in Fachpunkten ausgedrückt wird, ist nicht garantiert, daß die Fachpunkte genau der erreichten Prozentzahl von Punkten entsprechen.

Punkte-Datenbank im WWW

- Fehler beim Notieren der Punkte für Hausaufgaben und Klausuren kommen vor.
- Deswegen haben Sie in diesem Kurs die Möglichkeit, Ihren Punktestand im WWW zu kontrollieren.
- Sie müssen sich in die Datenbank eintragen (und ein Passwort vergeben) bevor die ersten Hausaufgabenpunkte eingetragen werden.

  Sonst kann sich jemand anders unter Ihrem Namen registrieren.

- Falls Datenschutz-Bedenken: Bitte melden.

Lehrbücher (1)

DB-Lehrbücher mit Teil über DBMS-Implementierung:

- Alfons Kemper, André Eickler:
  Datenbanksysteme. Eine Einführung. 7. Auflage.

  Oldenbourg, 2009, ISBN: 3486590189, 718 Seiten.

- Ramez Elmasri, Shamkant B. Navathe:
  Fundamentals of Database Systems, 5th Edition.

  Addison Wesley, 2006, ISBN 032141506X, 1168 pages. Deutsche Ausgabe ist um die für die Vorlesung interessanten Teile gekürzt.

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

  McGraw-Hill, 2002, ISBN: 0071151109, 1065 pages.

Lehrbücher (2)

Lehrbücher speziell über DBMS-Implementierung:

- Gunter Saake, Andreas Heuer:
  Datenbanken: Implementierungstechniken, 2. Aufl.

  MITP-Verlag, 2005, ISBN 3-8266-1438-0, 869 Seiten.

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

  Prentice Hall, 2008, ISBN: 0131354280, 1248 pages.
  Nur der zweite Teil ist über DBMS-Implementierung. Früher waren es zwei Bände. Der zweite heißt "Database System Implementation". Eventuell noch erhältlich, aber man vergleiche Preis.

Lehrbücher (3)

Lehrbücher über DBMS-Implementierung, Forts.:

- Theo Härder, Erhard Rahm: Datenbanksysteme: Konzepte und Techniken der Implementierung.

  2. Aufl., Springer, 2001, ISBN 3-540-42133-5, 582 pages, 40 Euro.

- Jim Gray, Andreas Reuter: Transaction Processing: Concepts and Techniques.

  Morgan Kaufmann, 1993, ISBN 1-55860-190-2, 1070 pages, 96 Euro.

Lehrbücher (4)

Oracle:

- Mark Gurry, Peter Corrigan: Oracle Performance Tuning, 2nd Edition.

  O'Reilly, 1996, ISBN 1-56592-237-9, 942 pages, 50 Euro.

- John Watson, Roopesh Ramklass, Bob Bryla: OCA/OCP Oracle Database 11g All-in-One Exam Guide: Exams 1Z0-051, 1Z0-052, 1Z0-053.

  McGraw-Hill/Oracle, 2009, ISBN: 0071629181, 1050 pages, CD.

- Oracle Documentation ist online verfügbar

  [http://www.oracle.com/technology/documentation/database.html]
  Empfehlenswert z.B. Oracle Concepts Manual.

Lehrbücher (5)

- Raul Chong, Xiaomei Wang, Michael Dang, Dwaine R. Snow: Understanding DB2: Learning Visually with Examples, 2nd Ed.

  IBM Press, 2007, ISBN: 0131580183, 1056 Seiten, 50.99 Euro.

- DB2 Handbücher sind online

  [http://www-01.ibm.com/support/docview.wss?rs=71&uid=swg27009552]
  [http://publib.boulder.ibm.com/infocenter/db2luw/v9/index.jsp]

- Tutorials zur DB2 Zertifizierung sind online

  [http://www.ibm.com/certify/tests/edu730.shtml]

Software (Oracle)

- In der Übung: Oracle (mit DBA-Rechten).

  Oracle ist ein Anbieter für relationale DBMS (bis vor kurzem knapp vor IBM Marktführer, IBM ist jetzt durch den Kauf von Informix in Führung gegangen). Der Dozent ist Oracle Certified Professional (Oracle 8 DBA). Diese Vorlesung ist daher etwas Oracle-lastig. Ich begrüsse es aber ausdrücklich, wenn Sie andere Systeme ausprobieren und mir über Gemeinsamkeiten und Unterschiede berichten. Das Institut wird für besonders interessierte Studierende auch noch weitere DBMS (IBM DB2, MS SQL Server, ...) zur Verfügung stellen.

- Durch die OAI-Teilnahme des Instituts können Sie bestimmte Oracle-Software kostenlos bekommen.

  Sie bekommen auch Vergünstigungen für die Zertifizierungsprüfungen (25%) und für die Beispielfragen von "Self Test Software" (50%).

Zertifizierung: Was ist das?

- Prüfung der (systemspezifischen) Kenntnisse durch den Datenbank-Hersteller.
- Im Erfolgsfall bekommt man
  - ◇ eine Urkunde,
  - ◇ das Recht, ein Logo auf seinen Visitenkarten und anderen Schriftstücken zu benutzen,
  - ◇ ggf. Zugriff auf eine spezielle Webseite.
- Die Zertifizierung gilt für eine bestimmte Version.

  Nach einiger Zeit sinkt der Wert der Zertifizierung, oder sie kann sogar durch den Anbieter als ausgelaufen (nicht mehr gültig) erklärt werden (z.B. bei Microsoft). Man braucht dann eine Upgrade-Prüfung.

Vorteile für Sie

- Die Zertifizierung belegt Detailkenntnisse von dem jeweiligen System, die in Vorlesungen normalerweise nicht vermittelt werden.
- Bei manchen potentiellen Arbeitgebern bringt die Zertifizierung entscheidene Pluspunkte für eine Bewerbung, bei anderen eher nicht.

  Falls der Arbeitgeber annimmt, daß ein Universitäts-Studium nur ganz theoretisches Wissen vermittelt, wird eine Zertifizierung einen wichtigen Unterschied machen, und eventuell etwas Berufspraxis aufwiegen, die Ihnen fehlt. Wenn der Arbeitgeber aber nur Persönlichkeit und Grundlagenwissen voraussetzt, und annimmt, daß Sie sich die Details bei ihm aneignen, bringt die Zertifizierung weniger: Größere Firmen können es sich leisten, ihre Mitarbeiter ggf. auch länger zu schulen.

Vorteile für den Arbeitgeber

- IT-Dienstleister und Unternehmensberatungen werben mit der Zertifizierung ihrer Mitarbeiter.
- Die Datenbank-Hersteller suggerieren, daß man die Zertifizierung nur mit viel praktischer Erfahrung am System bekommen kann, und daß dies dann auch zu sicherer und effektiverer Arbeit führt.

  Als Administrator kann man viel kaputt machen. Es ist für die Firma natürlich wichtig, zu wissen, ob sie Ihren Fachkenntnissen vertrauen kann. Eine Zertifizierung ist sicher ein Hinweis. Man kann eine Zertifizierungsprüfung aber auch ohne allzu viel praktische Erfahrung bestehen. Wenn man es nur durch "Stoff pauken" geschafft hat, vergißt man die Details häufig schnell wieder.

Vorteile für DBMS-Anbieter

- Schulung und Zertifizierung ist ein Geschäft.

  DBMS-Anbieter verdienen keineswegs nur mit dem Verkauf ihrer Software. Die Zertifizierungsprüfungen kosten Geld, und häufig brauchen Kandidaten vorher Schulungen, Bücher, Beispielfragen, etc.
- Wer eine Zertifizierungsprüfung geschafft hat, wird seinem Arbeitgeber auch eher das entsprechende DBMS empfehlen.

  Aufgrund des Aufwands sind Doppel-Zertifizierungen selten.
- Eine Technologie kann sich nur durchsetzen, wenn es ausreichend viele geschulte Experten gibt.

  Anerkennung durch Zertifizierung eventuell Anreiz.

Nachteile für Sie

- Kostet viel Zeit (und eventuell auch Geld).
- Das erworbene Wissen ist einem wesentlichen Teil sehr systemspezifisch.

  Natürlich sind viele Konzepte z.B. in DB2 und in Oracle recht ähnlich. Genaue Kenntnis eines Systems könnte Ihnen dann helfen, sich leichter in das andere System einzuarbeiten (wenn die Unterschiede in den Details Sie nicht durcheinander bringen). Für die Prüfung werden Sie aber auch Dinge lernen müssen, die nur für das eine System gelten.
- Sie werden auch Dinge auswendig lernen müssen, die im realen Leben auch ein erfahrener DBA eher nachschlagen würde, weil er sie so selten braucht.

  Zu einem Teil kann man aber raten. Erfahrung verbessert Intuition.

Zertifizierung: Ablauf (1)

- Ankreuzaufgaben am Rechner
  - ◇ Z.B. 60 Fragen in 90 Minuten
  - ◇ Typischerweise 4–5 mögliche Antworten

    Meist ist genau eine richtig. Es gibt aber auch Aufgaben, bei denen mehrere Antworten richtig sind, dann wird explizit gesagt, wie viele man ankreuzen soll. Die beiden Arten von Aufgaben unterscheiden sich auch durch Radiobuttons vs. Checkboxes.
  - ◇ Typischerweise muß man 60–65% der Fragen richtig beantwortet haben, um zu bestehen.

    Der genaue Prozentsatz, den man erreicht hat, steht (soweit mir bekannt) nicht auf dem Zeugnis.

Zertifizierung: Ablauf (2)

- Es sind keinerlei Unterlagen erlaubt (man muß alles auswendig wissen).

  Normalerweise bekommt man ein leeres Blatt Papier und einen Stift für Notizen während der Prüfung. Wenn die Aufsichtsperson streng ist, darf man das Papier hinterher nicht mitnehmen.
- Die Zeit ist knapp, aber nicht sehr knapp.

  Da keine Unterlagen erlaubt sind, kann man nicht mehr viel machen, wenn man die Antwort nicht weiß. Es lohnt sich dann nicht, zu lange hin und her zu überlegen. Bei meiner Oracle8-Zertifizierung bin ich nie in zeitliche Bedrängnis gekommen. Bei meiner DB2-Zertifizierung ein bißchen. Bei einem Microsoft-Probetest, den ich mal probiert habe, war es hoffnungslos. Das korrespondiert allerdings mit meinem Wissen der verschiedenen Systeme.

Oracle-Zertifizierung

- Oracle 11g Certified Associate
  - ◇ 1Z0-051 Oracle Database 11g: SQL Fundamentals I

    Online (von zu Hause aus, Kosten: 99 €, -40% OAI Rabatt). Es gibt Alternativen: “1Z0-007 Introduction to Oracle9i SQL” oder “1Z0-047 Oracle Database SQL Expert” (gibt alleine Titel).
  - ◇ 1Z0-052 Oracle Database 11g: Administration I

    Diese Prüfung muß unter Aufsicht abgelegt werden (bei Schulungsunternehmen in Halle, Kosten: 154 €, -40%).
- Oracle 11g DBA Certified Professional

  Weitere Prüfung und Nachweis von Schulung. [http://education.oracle.com/certification/]

IBM DB Zertifizierungen

- IBM Certified Database Associate
  DB2 9 Fundamentals

  Test 730 (DB2 9 Family Fundamentals)

- IBM Certified Database Administrator
  DB2 9 DBA for Linux UNIX and Windows

  (Test 700, 730, 914, 915, 917) und Test 731
  oder 700, 701 und 736 (Upgrade Test)

- IBM Certified Advanced Database Administrator
  DB2 9 for Linux UNIX and Windows

  (Test 700 oder 730) und (Test 731 oder 736) und Test 734
  [http://www-03.ibm.com/certify/certs/dm_index.shtml]

Test 730 (1)

- 64 Fragen, 90 Minuten, mindestens 59%
- Planning (14%)

  Restricting data access, Features/functions available in DB2 tools (e.g., control center, configuration advisor, configuration assistant, command line processor), Database Workloads (OLTP vs. warehousing), non-relational data concepts (extenders), XML data

- Security (11%)

  DB2 products, privileges and authorities, encryption options, identify results of given DDL statement (grant/revoke/connect).

- Working with Databases and DB Objects (17%)

  Identify and connect to DB2 servers and databases, identify DB2 objects, ability to create objects (SQL DDL)

Test 730 (2)

- Working with DB2 Data using SQL (23.5%)

  Identify results of given SQL DML statement (SELECT, SORT, GROUP, UPDATE, DELETE, INSERT), Knowledge of transactions (COMMIT/ROLLBACK, transaction boundaries), call a procedure or user defined function, identify results of given XQuery statement.

- Working with DB2 Tables, Views, Indexes (23.5%)

  DB2 data types, CREATE TABLE, referential integrity, data constraints, characteristics of a table/view/index, identify when triggers should be used, data type operations for storing XML data

- Data Concurrency (11%)

  Identify factors that influence locking, list objects on which locks can be obtained, identify characteristics of DB2 locks, identify the isolation level that should be used for a given situation

Ausstattung der DB-Gruppe

- Sun Fire V440 (4 CPUs, 16 GB RAM)

  ca. 14.500 Euro (2005)

- Sun Fire V490 (4 Doppelkern-CPUs, 16 GB RAM)

  ca. 30.000 Euro (2005)

- Eine kleinere Sun, zwei Linux-Server
- SAN (Storage Area Network) mit
  - ◇ Fibre Channel RAID System: 36×146 GB = 5.2 TB

    ca. 36.000 Euro (2005)

  - ◇ SATA RAID System: 12×400 GB = 4.8 TB

    ca. 13.000 Euro (2005)

Verbesserung der Lehre

- Gute Lehre ist für mich wichtig.
- Vorschläge zur Verbesserung der Vorlesung sind sehr willkommen. Fragen sind sehr willkommen.

  Auch Korrekturen zu den Folien oder nützliche Links für die Webseite.
- Vorlesung soll nicht Monolog werden. Ich frage gelegentlich einzelne Studierende, aber Sie dürfen die Frage weitergeben ("schieben").
- Bei großem Interesse: Zertifizierungs-AG. Erfordert aber aktive Mitarbeit (z.B. Vorträge). Keine LP.