Dr. Stefan Brass
Institut für Informatik
Martin-Luther-Universität Halle-Wittenberg

11. Juli 2003

Vorlesung “Datenbanken II”
— Endklausur —

Name: ____________________

Matrikelnummer: ____________________

Aufgabe	Punkte	Max. Punktzahl
1 (Semantische Fehler in SQL-Anfragen)		4
2 (Data Dictionary)		4
3 (Platten)		4
4 (Heap-Dateien)		5
5 (Indexe)		4
Summe		21

Anleitung

- Sie haben 60 Minuten Zeit für diese Klausur. Wir werden ca. um 8^{30} beginnen, so daß der Abgabetermin ca. um 9^{30} ist (genauer Termin wird an die Tafel geschrieben). Sie dürfen erst ca. 10 Minuten vor Ende der Klausur gehen, da danach niemand mehr auf Toilette gehen kann (Es darf immer nur ein Teilnehmer gleichzeitig den Raum verlassen.)
- Bitte schreiben Sie die Antworten in den vorgegebenen Platz. Falls der Platz nicht ausreicht, können Sie weitere Blätter oder die Rückseite verwenden, aber bitte markieren Sie klar, daß es noch eine Fortsetzung gibt. Schreiben Sie Ihren Namen und die Nummer der Aufgabe auf alle zusätzlichen Blätter.
- Bitte fragen Sie, wenn Sie die Aufgabe nicht verstehen, oder die Aufgabe Ihnen mehrdeutig erscheint.
- Die Klausur hat 10 Seiten. Bitte prüfen Sie die Vollständigkeit.
- Bitte nehmen Sie keine Klausur mit und auch keine Notizen über die Aufgaben. Sie bekommen die Klausur zurück, sobald sie korrigiert ist.
- Sie können alle schriftlichen Aufzeichnungen verwenden, die vor Beginn der Klausur entstanden sind. Dagegen sind Computer (Laptops etc.) leider nicht zulässig. Es ist auch nicht viel Zeit für langes Suchen in den Aufzeichnungen.
- Wenn Sie unbedingt einen Taschenrechner verwenden wollen, dürfen Sie nur sehr einfache benutzen, und müssen sie zur Kontrolle kurz vorzeigen. Geräte, auf denen eventuell Oracle Lite laufen könnte, oder die Funkverbindungen erlauben, sind nicht zugelassen. Die Aufgaben sollten sich leicht im Kopf rechnen lassen.
- Bitte vermeiden Sie jedes Verhalten, das als Täuschungsversuch mißverstanden werden könnte. Achten Sie auf einen ausreichenden Abstand zu Ihren Nachbarn und setzen Sie sich nicht direkt neben die Teilnehmer, mit denen Sie vorher immer die Hausaufgaben zusammen abgegeben haben. Schauen Sie nicht in Richtung der Klausuren anderer Teilnehmer und sprechen Sie nicht mit anderen Teilnehmern, bis alle Klausuren abgegeben sind. Tauschen Sie keinesfalls irgendwelche Papiere oder Gegenstände aus. Notfalls rufen Sie bitte den Dozenten zur Kontrolle.
- Bitte stellen Sie sicher, daß ich Ihre Handschrift lesen kann. Verwenden Sie keine rote Tinte. Sie dürfen mit Bleistift schreiben, aber es könnte bei späterer Kritik an meiner Korrektur eventuell schwierig werden, zu beweisen, daß Sie etwas geschrieben haben. Dennoch sollten Sie nachfragen, wenn Sie meine Korrektur nicht verstehen.
- Meine Zeitschätzung ist: Aufgabe 1 je 3 Minuten (d.h. 12 min gesamt), Aufgabe 2 je 5 min (d.h. 10 min gesamt), Aufgabe 3 je 2 min (d.h. 8 min gesamt), Aufgabe 4 12 min, Aufgabe 5 10 min. die Zeitreserve beträgt dann noch 8 Minuten. Falls Sie Schwierigkeiten mit einer Aufgabe haben, bearbeiten Sie zunächst die anderen. Es empfiehlt sich aber, alle Aufgaben wenigstens zu versuchen (wenn Sie nichts hinschreiben, bekommen Sie für die Aufgabe sicher 0 Punkte).

Beispiel-Datenbank

Gegeben sei eine Datenbank zur Abspeicherung der Verkaufsdaten (Kassenbons) in einem Supermarkt.

- Die erste Tabelle enthält die Daten der Waren (Produkte, die in dem Supermarkt verkauft werden):

WARE			
WID	NAME	MENGE	PREIS
1001	Milch	1 Liter	0.79
1002	Mehl	1 kg	1.00
1003	Eier	6 Stück	1.50
⋮	⋮	⋮	⋮

  `WID` ist Schlüssel dieser Tabelle.

- Die zweite Tabelle enthält folgende Daten über Kassenbons:

BON			
BNR	ZEIT	KASSE	KASSIERER
100001	10.07.03 17:05	1	Schmidt
100002	10.07.03 17:07	1	Schmidt
100003	10.07.03 17:08	2	Meier
⋮	⋮	⋮	⋮

  `BNR` ist Schlüssel dieser Tabelle.

- Die dritte Tabelle enthält die einzelnen Positionen der Kassenbons:

BPOS			
BNR	POS	WID	ANZAHL
100001	1	1002	1
100001	2	1003	1
100001	3	1002	1
100002	1	1003	3
⋮	⋮	⋮	⋮

  Die Spalte `POS` ist eine laufende Nummer, mit der die Positionen eines Bons identifiziert werden (gewissermaßen die Zeilennummer des Bons). Daher ist die Kombination von `BNR` und `POS` Schlüssel dieser Tabelle. `BNR` ist ein Fremdschlüssel, der auf `BON` verweist, und `WID` ist ein Fremdschlüssel, der auf `WARE` verweist. Es ist möglich, daß die gleiche Ware auf dem gleichen Bon mehrfach auftaucht (wenn mehrere Exemplare gekauft wurden und sie nicht direkt hintereinander auf dem Band lagen).

Aufgabe 1 (Semantische Fehler in SQL-Anfragen) 4 Punkte

Betrachten Sie die folgenden SQL-Anfragen an die Beispiel-Datenbank. Manche dieser Anfragen enthalten Fehler oder unnötige Komplikationen. Bitte kreuzen Sie die jeweils richtige Aussage an.

a) Folgende Anfrage soll Kassenbon-Nummern liefern, bei denen Milch verkauft wurde:

```
SELECT DISTINCT B.BNR
FROM   WARE W, BPOS P, BON B
WHERE  W.WID = P.WID AND P.BNR = B.BNR
AND    W.NAME = 'Milch'
```

Was meinen Sie zu dieser Anfrage?

- ☐ DISTINCT ist überflüssig.
- ☐ Inkonsistente Bedingung.
- ☐ Unnötiger Join.
- ☐ Join-Bedingung fehlt.
- ☐ Sonstiges Problem: ____________________
- ☐ Perfekt.

b) Folgende Anfrage soll Waren (Produkte) liefern, die Käse in ihrem Namen enthalten (als Substring):

```
SELECT DISTINCT WID, NAME
FROM   WARE W
WHERE  W.NAME LIKE '%Käse%'
```

Wie schätzen Sie diese Anfrage ein?

- ☐ DISTINCT ist überflüssig.
- ☐ Inkonsistente Bedingung.
- ☐ Statt LIKE sollte man besser = schreiben.
- ☐ Sonstiges Problem: ____________________
- ☐ Perfekt.

c) Folgende Anfrage soll Kassenbon-Nummern liefern, bei denen die Ware 1002 mindestens zweimal verkauft wurde:

```
SELECT   BNR
FROM     BPOS
WHERE    WID = 1002
GROUP BY BNR, POS
HAVING   SUM(ANZAHL) >= 2
```

Ist diese Anfrage korrekt?

- ☐ Inkonsistente Bedingung.
- ☐ Falsches `GROUP BY`.
- ☐ Man sollte die `HAVING`-Bedingung unter `WHERE` aufschreiben.
- ☐ Sonstiges Problem: ____
- ☐ Perfekt.

d) Diese Anfrage soll Kassenbon-Nummern liefern, bei denen sowohl Mehl als auch Eier verkauft wurden:

```
SELECT DISTINCT P.BNR
FROM   BPOS P, WARE W1, WARE W2
WHERE  P.WID = W1.WID AND W1.NAME = 'Mehl'
AND    P.WID = W2.WID AND W2.NAME = 'Eier'
```

Kreuzen Sie die beste Erklärung zu dieser Anfrage an:

- ☐ Inkonsistente Bedingung.
- ☐ Unnötiger Join.
- ☐ Join-Bedingung fehlt.
- ☐ Sonstiges Problem: ____
- ☐ Perfekt.

Aufgabe 2 (Data Dictionary) 4 Punkte

Formulieren Sie die folgenden Anfragen an das Oracle Data Dictionary in SQL. Es könnten u.a. diese Tabellen nützlich sein:

- `DICT(TABLE_NAME, COMMENTS)`
- `DICT_COLUMNS(TABLE_NAME, COLUMN_NAME, COMMENTS)`
- `USER_CATALOG(TABLE_NAME, TABLE_TYPE)`
- `ALL_CATALOG(OWNER, TABLE_NAME, TABLE_TYPE)`
- `TABS(TABLE_NAME, TABLESPACE_NAME, ...)`
- `USER_COLUMNS(TABLE_NAME, COLUMN_NAME, DATA_TYPE, ..., COLUMN_ID, ...)`
- `USER_CONSTRAINTS(OWNER, CONSTRAINT_NAME, CONSTRAINT_TYPE, TABLE_NAME, R_OWNER, R_CONSTRAINT_NAME, ...)` `CONSTRAINT_TYPE` ist `P` für Primärschlüssel.
- `USER_CONS_COLUMNS(OWNER, CONSTRAINT_NAME, TABLE_NAME, COLUMN_NAME, POSITION)`

Selbstverständlich können Sie auch andere Tabellen des Oracle Data Dictionary verwenden.

a) Geben Sie alle Tabellen aus, die Ihnen gehören, und eine Spalte mit dem Namen “`WID`” haben. Es sollen nur echte Tabellen berücksichtigt werden, keine Sichten.

b) Auf welchen von Ihren Tabellen ist kein Primärschlüssel deklariert?

Aufgabe 3 (Platten) 4 Punkte

Kreuzen Sie die jeweils richtige bzw. beste Antwort an. Machen Sie pro Unteraufgabe nur ein Kreuz. Pro Unteraufgabe gibt es einen Punkt. Falls Sie die Lösung nicht wissen, raten Sie (wenn Sie nichts ankreuzen, haben Sie den Punkt auf jeden Fall verloren).

a) Nehmen wir an, die Tabelle `BPOS` benötigt 100 MB Speicherplatz. In welcher Zeit kann Sie von einer normalen Platte gelesen werden (unter der Voraussetzung, daß sie "an einem Stück" abgespeichert ist, nicht beliebig über die Platte verteilt)? Kreuzen Sie an, was der richtigen Antwort am nächsten kommt.

- ☐ Ca. 20 μs (Microsekunden).
- ☐ Ca. 10 ms (Millisekunden).
- ☐ Ca. 5 Sekunden.
- ☐ Ca. eine Minute.
- ☐ Ca. 5 Minuten.

b) Welche der folgenden Zeiten hängt von der Entfernung des Kopfes zum gewünschten Cylinder (Spur) ab?

- ☐ Seek Zeit.
- ☐ Latenzzeit.
- ☐ Transferzeit.

c) Angenommen, Sie haben 10 Platten zu je 10 GB. Welchen Plattenplatz können Sie maximal für Nutzdaten haben, wenn Sie RAID-5 (Parity) verwenden?

- ☐ 200 GB.
- ☐ 100 GB.
- ☐ 90 GB.
- ☐ 50 GB.
- ☐ 10 GB.

d) Gibt es eventuell Gründe dafür, den maximal möglichen Plattenplatz bei RAID-5 nach c) nicht auszunutzen? Kreuzen Sie die beste Antwort an.

- ☐ Nein. Man sollte immer den maximal möglichen Plattenplatz wählen.
- ☐ Erhöhte Sicherheit. RAID-5 kann nicht immer nach dem Defekt einer einzelnen Platte die Daten rekonstruieren.
- ☐ Performance-Einbruch bei einem Plattenfehler.

Aufgabe 4 (Heap-Dateien) 2+1+2=5 Punkte

Diese Aufgabe bezieht sich auf Oracle.

a) Wie lang werden Tupel in der Relation `BPOS` normalerweise sein (Anzahl Bytes, `AVG_ROW_LEN`)? Die Tabelle hat vier Spalten, sämtlich numerisch:

- Das Attribut `BNR` ist als `NUMERIC(6)` deklariert und benötigt auch alle 6 Stellen.
- Das Attribute `POS` ist als `NUMERIC(3)` deklariert, aber fast immer wird nur eine Stelle benötig.
- Das gleiche gilt für das Attribut `ANZAHL`.
- `WID` ist als `NUMERIC(4)` deklariert und die Daten haben auch immer vier Stellen.

Bitte geben Sie nicht nur das Ergebnis an, sondern nennen Sie auch kurz die verschiedenen Komponenten.

b) Auf dieser Tabelle gibt es keine nachträglichen Updates, nur Einfügungen. Wie würden Sie `PCTFREE` dann wählen?

- ☐ Klein, z.B. 1%.
- ☐ Mittel, z.B. 10%.
- ☐ Groß, z.B. 50%.

c) Unabhängig von b) sei angenommen, daß `PCTFREE=10` gewählt wurde (das ist der Default). Angenommen, Blöcke sind 200 Byte groß. Wie viele Zeilen passen dann in einen Block? Die Blockgröße ist natürlich unrealistisch klein, aber so sollten Sie ohne Taschenrechner auskommen. Geben Sie bitte wieder nicht nur das Endergebnis an, sondern auch, wie Sie es berechnet haben.

Aufgabe 5 (Indexe) 4 Punkte

Kreuzen Sie die jeweils richtige bzw. beste Antwort an. Machen Sie pro Unteraufgabe nur ein Kreuz. Pro Unteraufgabe gibt es einen Punkt. Falls Sie die Lösung nicht wissen, raten Sie (wenn Sie nichts ankreuzen, haben Sie den Punkt auf jeden Fall verloren).

a) Folgende Anfrage sei gegeben:

```
SELECT COUNT(*)
FROM   BPOS
WHERE  WID = 1003
```

Das Ergebnis sei 10. Aufgrund der sequentiellen Einfügungen kann man davon ausgehen, daß die Tupel für eine Ware auf verschiedene Blöcke verstreut sind. Wenn es einen Index über der Spalte `WID` gibt, wie viele Blockzugriffe wird die Anfrage benötigen? Der Index habe die Höhe 4. [Dies ist die einzige Aufgabe der Klausur, die versucht, Sie reinzulegen.]

- ☐ Ca. 4.
- ☐ Ca. 10.
- ☐ Ca. 14.
- ☐ Ca. 40.

b) Es kommt öfters vor, daß die Positionen eines Bons ausgedruckt werden müssen:

```
SELECT   POS, WID, ANZAHL
FROM     BPOS
WHERE    BNR = 100001
ORDER BY POS
```

Da `BNR` und `POS` zusammen ein Schlüssel sind, gibt es einen Index über der Spaltenkombination. Spielt die Reihenfolge dieser Spalten in der Index-Deklaration eine Rolle (für den Schlüssel wäre es ja egal)? Kann der Index überhaupt verwendet werden, oder sollte man einen Index nur über `BNR` anlegen?

- ☐ Ein Index über (`BNR`, `POS`) ist gut für diese Anfrage. Ein Index über (`POS`, `BNR`) könnte dagegen nicht verwendet werden.
- ☐ Gerade umgekehrt.
- ☐ Die Reihenfolge der Spalten spielt keine Rolle, beide Indexe können verwendet werden.
- ☐ Die Indexe können beide nicht verwendet werden, man sollte einen Index nur über `BNR` alleine anlegen.

c) Kann es vorkommen, daß eine Anfrage mit Index ausgewertet länger läuft als ein Full Table Scan?

- ☐ Ja. Es kann vorkommen, daß ein Full Table Scan schneller ist als die Auswertung mit Index.
- ☐ Das ist unmöglich. Ein Full Table Scan ist immer die schlechteste Möglichkeit. Er sollte nur verwendet werden, wenn kein Index anwendbar ist.

d) Gesucht sind die verschiedenen Sorten "Du darfst" Lebensmittel (z.B. "Du darfst Käseaufschnitt", "Du darfst Fleischsalat") mit folgender Anfrage:

```
SELECT NAME
FROM   WARE
WHERE  NAME LIKE 'Du darfst %'
```

Wäre ein B-Baum-Index über `WARE(NAME)` für diese Anfrage nützlich?

- ☐ Ja.
- ☐ Nein.