

Prof. Dr. Nikolaus Augsten

Jakob-Haringer-Str. 2
5020 Salzburg, Austria
Telefon: +43 662 8044 6347
E-Mail: nikolaus.augsten@sbg.ac.at



Datenbanken II
Datenbanken Vertiefung
Wintersemester 2014/2015

Prüfung

10.06.2015

Name: _____ Matrikelnummer: _____

Hinweise

- Bitte überprüfen Sie die Vollständigkeit des Prüfungsbogens (6 nummerierte Seiten).
- Schreiben Sie Ihren Namen und Ihre Matrikelnummer auf jedes Blatt des Prüfungsbogens und geben Sie alle Blätter ab.
- Grundsätzlich sollten Sie alle Antworten auf den Prüfungsbogen schreiben.
- Sollten Sie mehr Platz für eine Antwort benötigen, bitte einen klaren Verweis neben die Frage auf die Seitennummer des zusätzlichen Blattes setzen.
- Keinen Bleistift verwenden. Keinen roten Stift verwenden.
- Verwenden Sie die Notation und die Lösungsansätze, die während der Vorlesung besprochen wurden.
- Aufgaben mit mehr als einer Lösung werden nicht bewertet.
- Als Unterlage ist ein beliebig (auch beidseitig) beschriftetes A4-Blatt erlaubt.
- Zeit für die Prüfung: 90 Minuten

Unterschrift _____

Korrekturabschnitt

Bitte frei lassen

Aufgabe	1	2	3	4	5	Summe
Maximale Punkte	12	14	12	24	18	80
Erreichte Punkte						

Aufgabe 1**12 Punkte**

Für ein Datenbanksystem sollen Slotted Pages der Größe 2048 Bytes (0...2047) dimensioniert werden.

Es wird Byte-Adressierung verwendet. Jedes Feld im Kopfteil und jeder Datensatz kann an einer beliebigen Byte-Adresse beginnen.

- 1.1 Ergänzen Sie in der folgenden Slotted Page die fehlenden Werte. Tupel A ist 127 Bytes groß, Tupel B ist 148 Bytes groß und Tupel C ist 93 Bytes groß. (9 Punkte)

									A	B	C
--	--	--	--	--	--	--	--	--	---	---	---

- 1.2 Was sind die Größen von allen Felder? (3 Punkte)

Name:

Matrikelnummer:

Aufgabe 2

14 Punkte

Eine Relation $R(A, B)$ ist in einer sequentiellen Datei nach A sortiert gespeichert (die Attribute A and B enthalten positive numerische Werte). Relation R enthält 10.000 Tupel. Jeder Wert von A kommt 50 mal vor und jeder Wert von B kommt 25 mal vor.

Block Größe $B = 2048$ Bytes, Tupel Größe $t = 64$ Bytes.

- 2.1 Versuchen Sie die folgende Anfrage möglichst effizient zu beantworten und schätzen Sie die maximale Anzahl der nötigen Blockzugriffe ab. (6 Punkte)

$$\sigma_{A=50 \wedge B=1000}(R)$$

- 2.2 Betrachten Sie folgende Festplatte: Sektorgröße = Block Größe = B , Sektoren/Spur S , mittlerer Spurwechsel sp ms, Drehzahl dz rpm, keine Interblock Gap.

Die ganze Relation R wird auf einer Spur gespeichert. Ein Block passt in den Puffer.

Schätzen Sie die Zeit für das Lesen der nötigen Blöcke von Aufgabe 2.1. (8 Punkte)

Betrachten Sie die folgende Relation:

Tupel ID	Name	Salary	Age
1	Alice	1000	25
2	Amelia	1600	28
3	Antonio	1200	17
4	Dante	1600	27
5	Elmo	1800	24
6	Gabriella	1600	23
7	Lucia	1200	21
8	Orso	1000	26
9	Raul	1600	30
10	Sonia	1400	18
11	Walter	1800	19

- 3.1 Erstelle und zeichne einen B+-Baum auf dem Attribut Age. Pro Knoten gibt es drei Zeiger. (3 Punkte)
- 3.2 Löschen Sie in lexikographischer Ordnung alle Tupel die Salary=1600 haben, und stellen Sie den resultierenden B+-Baum dar. (5 Punkte)
- 3.3 Erstelle und zeichne einen Secondary Index (ISAM, einstufig) auf dem Attribut Salary. Für die Duplikate sollen Buckets verwendet werden (ein Bucket hat Platz für drei Pointer). Für jeden Datenblock gibt es einen eigenen Indexeintrag. (4 Punkte)

Name:

Matrikelnummer:

Aufgabe 4

24 Punkte

Die Relation $R[\underline{A}, B, D, E]$ hat $n_R = 20000$ Tupel auf $b_R = 4000$ Blöcken, die Relation $S[\underline{B}, C]$ hat $n_S = 2000$ Tupel auf $b_S = 500$ Blöcken.

- 4.1 Schätzen Sie die Zahl der Block-Zugriffe für einen natürlichen Index-Nested Loop Join ab (Worst-Case Abschätzung). Als Index wird ein Sekundärindex auf $R.B$ verwendet. Der Index ist als B^+ -Baum mit 20 Suchschlüsseln pro Knoten organisiert. Die Schlüssel in den Blättern sind eindeutig. Die Duplikate werden in Buckets verwaltet. (14 Punkte)

Annahme: jedes Bucket passt in einen Block und hat Platz für 10 Pointer, jeder Wert von B in R kommt 25 mal vor.

- 4.2 Was ist die beste Auswertungsstrategie für folgende Selektion, wenn es einen B^+ -Baum Sekundärindex auf (A, D) der Relation R gibt? Listen Sie die nötigen Schritte auf. (10 Punkte)

$$\sigma_{A \geq 100 \wedge B = 123 \wedge D = 'b' \wedge E = 't'}(R)$$

Betrachte die folgenden Relationen und eine SQL-Anfrage.

(B)boats(bid, name, colour)

(S)ailors(sid, name, rating, age)

(R)eservations(bid, sid, day)

```
SELECT B.name
FROM Boats B, Sailors S, Reservations R
WHERE B.bid = R.bid
      AND S.bid = R.bid
      AND S.age >= 40
      AND B.colour = 'red';
```

- 5.1 Zeichne die algebraische Normalform von der gegebenen SQL-Anfrage als Operatorbaum (Projektionen-Selektionen-kartesische Produkte). *(6 Punkte)*
- 5.2 Wende günstige Äquivalenzregeln an, um die Anfrage zu optimieren (heuristische Optimierung). *(12 Punkte)*