

# Datenbanksysteme I

## 8. Übungsblatt, Sommersemester 2005

Besprechung am Mittwoch, den 15.06.2005

1. Gegeben sei ein Relationenschema  $\mathcal{R}$ (Flugnummer, Wochentag, Flugzeugtyp) für Flugverbindungen.

Betrachten Sie die Relationen  $R_1$  und  $R_2$  in Abbildung 1 und Abbildung 2. Ein Tupel der Relation gibt an, an welchem Wochentag ein Flug stattfindet und welcher Flugzeugtyp dafür eingesetzt werden kann.

$R_1$		
Flugnummer	Wochentag	Flugzeugtyp
106	Montag	747
106	Donnerstag	747
106	Montag	DC10
106	Donnerstag	DC10
204	Montag	747
204	Donnerstag	747

Abbildung 1: Relation  $R_1$

- (a) Welche funktionalen Abhängigkeiten erfüllt  $R_1$ ?
- (b) Gilt  $R_1 = \pi_{Flugnummer, Wochentag}(R_1) \bowtie \pi_{Flugnummer, Flugzeugtyp}(R_1)$ ?
- (c) Welche funktionalen Abhängigkeiten erfüllt  $R_2$ ?
- (d) Gilt hier  $R_2 = \pi_{Flugnummer, Wochentag}(R_2) \bowtie \pi_{Flugnummer, Flugzeugtyp}(R_2)$ ?
- (e) Wieso gilt für ein  $i \in \{1, 2\}$   $R_i = \pi_{Flugnummer, Wochentag}(R_i) \bowtie \pi_{Flugnummer, Flugzeugtyp}(R_i)$ , für das andere jedoch nicht?
2. Für die Relation  $R$ , die auf 182 hintereinanderliegenden Blöcken auf der Platte gespeichert ist, wurde ein geballter  $B^+$ -Baum (clustered  $B^+$ -tree) auf dem Schlüsselattribut  $A$  mit der Höhe 2 angelegt (siehe Abbildung 3).

Für einen Zugriff auf einen bestimmten Block auf einer Festplatte werden drei Arbeitsvorgänge benötigt. Zunächst muß der Schreib-/Lesekopf auf die entsprechende Spur plaziert werden. Die dazu benötigte Zeit wird als *Seek Time* ( $t_s$ ) bezeichnet.

$R_2$		
Flugnummer	Wochentag	Flugzeugtyp
106	Montag	747
106	Donnerstag	747
106	Donnerstag	DC10
204	Montag	747
204	Donnerstag	747

Abbildung 2: Relation  $R_2$

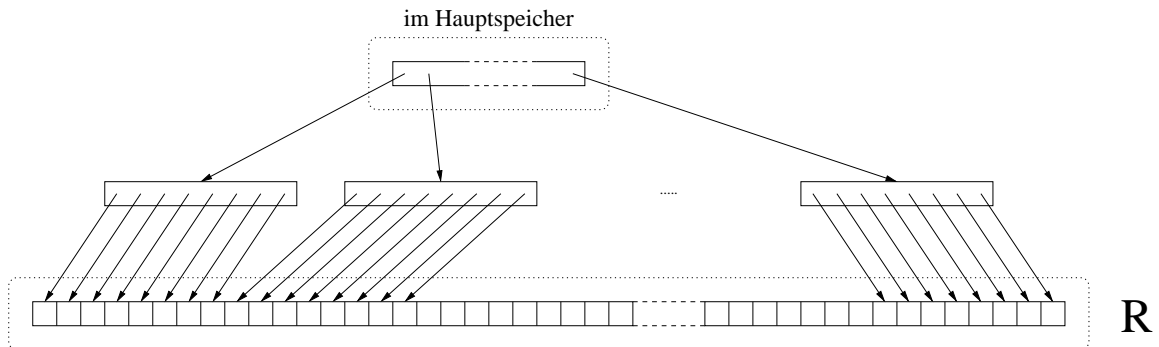


Abbildung 3: geclusterter B<sup>+</sup>-Baum

Dann wird gewartet, bis sich durch die Rotation der Platte der gesuchte Block am Kopf vorbeibewegt. Die zu erwartende Verzögerung wird *Latenzzeit* ( $t_l$ ) genannt. Im dritten Schritt wird der Block gelesen (*Lesezeit*,  $t_t$ ). Für diese Aufgabe gehen Sie von folgenden Daten für die Platte aus:

Seek Time ( $t_s$ )	12ms
Latenzzeit ( $t_l$ )	6ms
Lese- oder Transferzeit ( $t_t$ )	2ms

Bei einem sequentiellen Scan durch die Relation muß einmal gesucht werden, einmal die Latenzzeit abgewartet werden und 182 Blöcke in den Hauptspeicher transferiert werden.

(a) Es werden Anfragen folgender Form an die Datenbank gestellt:

```
select *
from R
where A=23
or    A=19828
or    A=125
or    ...;
```

Ab wieviel verschiedenen gesuchten Werten für  $A$  ist ein sequentieller Scan genauso schnell wie die Zugriffe über den Index?

Gehen Sie davon aus, daß jede Seite des B<sup>+</sup>-Baums in einem Block liegt und zum Holen dieses Blocks das Suchen, die Latenz und die eigentliche Übertragung abgewartet werden muß. Aus Effizienzgründen wird der Wurzelknoten des B<sup>+</sup>-Baums im Hauptspeicher gehalten, muß also nicht von der Platte geholt werden. Gehen Sie weiterhin davon aus, daß jeder der gesuchten Werte für *A* auf einer anderen Blattseite liegt.

(b) Nun werden folgende Anfragen an die Datenbank gestellt:

```
select *
from R
where A > 23;
```

Bei dieser Art der Anfrage (Range-Query) kann nach dem Finden des Einstiegs- punktes in die Relation sequentiell weitergescannt werden. Wieviele Blöcke der Relation *R* müssen bei einer Range-Query durchlaufen werden, so daß der sequentielle Scan genauso schnell ist wie der Zugriff über den Index?

3. Die Datenelemente 23, 25, 15, 7, 5, 8, 9, 10, 11, 12, 13, 26, 27, 28, 1, 2, 3 sollen mit Hilfe des erweiterbaren Hashings indiziert werden. Nehmen Sie an, daß auf einer Seite des Datenbanksystems drei Datenelemente und beliebig viele Zeiger Platz finden. Fügen Sie die Elemente aus der obigen Liste in der angegebenen Reihenfolge in eine erweiterbare Hashtabelle ein. Zeichnen Sie dabei die Tabelle neu, falls sich signifikante Änderungen ergeben.

Nehmen Sie an, daß die verwendete Hashfunktion jede Zahl auf ihren (fünfstelligen) Binärcode abbildet.

4. (Elektronische Abgabe)

Gegen welche der folgenden mehrwertigen Abhängigkeiten verstößt die gegebene Ausprägung der Relation *R*?

R		
A	B	C
a1	b1	c1
a1	b2	c1
a2	b1	c2
a2	b2	c2
a3	b1	c1

(a) (1 Punkt)  
 $A \twoheadrightarrow B$

(b) (1 Punkt)  
 $A \twoheadrightarrow C$

(c) (1 Punkt)  
 $B \twoheadrightarrow A$

(d) (1 Punkt)  
 $C \twoheadrightarrow A$

(e) (1 Punkt)  
 $AB \twoheadrightarrow C$

5. (Elektronische Abgabe: jeweils 1 Punkt)

Welche der folgenden Aussagen sind wahr, welche sind falsch?

- Es kann nur einen geballten Primärindex auf einer Relation geben, aber beliebig viele Sekundärindexe.
- Der ursprünglich vergebene Tupel-Identifikator (TID) wird durch das Verschieben eines Tupels nicht geändert.
- Der Zugriff über einen Index ist immer schneller als der direkte Zugriff auf eine Relation.
- Ein Index erhöht im Allgemeinen zwar die Zugriffsgeschwindigkeit, verlangsamt aber die Update-Operationen.