UNIVERSITÄT DES SAARLANDES
Fachbereich 14– Informatik
Prof. Dr.–Ing. G. Weikum
Dipl.–Inform. A. Kraiß

Datenbanksysteme WS 1998/99

Zweite Teilklausur 22.01.1999

Aufgabe 1: Join-Index (2+8+4=14 Punkte)

Gegeben seien die Relationen *VERKAUF*(*PNr, KNr, Anzahl*) und *KUNDEN*(*KNr, Adresse*). Die unterstrichenen Attribute bilden den Primärschlüssel. Die Relation *VERKAUF* umfasse *10 Mio.* Tupel, wobei jedes Tupel eine konstante Größe von *100 Bytes* habe. Die Relation *KUNDEN* umfasse *1000* Tupel, wobei jedes Tupel eine konstante Größe von *20 Bytes* habe. Die nutzbare Nettogröße einer Seite sei *2000 Bytes*. Zur Speicherung eines Wertes des Attributes *KNr* benötige man *4 Bytes*. Jeder Kunde soll alle Produkte gekauft haben. Zur effizienten Ausführung eines Natural Joins zwischen den beiden Relationen existiere ein Join-Index in Form eines B*-Baumes. Zur Speicherung eines TIDs sowie zur Speicherung eines Sohnzeigers benötige man *4 Bytes*.

(a) Bestimmen Sie die Höhe *h* des Join-Index. Ein Index, der nur aus der Wurzel besteht, habe eine Höhe von *1*. Ignorieren Sie den Platzbedarf zum Verketten der Blattseiten.

(b) Nehmen Sie im folgenden an, daß neben den beiden Relationen lediglich der Join-Index aus (a), jedoch mit einer Höhe von *h=4*, existiere. Bestimmen Sie die Anzahl der Seitenzugriffe für die folgenden Join-Ausführungsstrategien zur Ausführung des Natural Join zwischen den Relationen *VERKAUF* und *KUNDEN*. Nehmen Sie an, daß den Joins ein Puffer zum Speichern von *2 Datenseiten* und *1 Indexseite* zur Verfügung steht, und daß jedes Verfolgen eines Sohnzeigers bzw. eines TIDs zu einem neuen Seitenzugriff führt.

  (i) Nested-Block-Join (ohne Index).

  (ii) Nested-Loop-Join mit *VERKAUF* in der äußeren Schleife und Indexzugriff in der inneren Schleife. Gehen Sie hierbei von einer für diesen Join günstigen Speicherungsreihenfolge der TIDs aus.

  (iii) Nested-Loop-Join mit *KUNDEN* in der äußeren Schleife und Indexzugriff in der inneren Schleife.

  (iv) Join durch Scan des Index.

(c) Sei *h* die Höhe des Join-Index mit *h>2*, und sei *p* die Anzahl der Blattseiten pro Schlüsselwert mit *p>1*. Die Anzahl der Tupel in der Relation *KUNDEN* sei durch t_K gegeben, und die Anzahl der Tupel in der Relation *VERKAUF* sei durch t_V gegeben. Weiterhin sei die Anzahl der Seiten für die Relation *KUNDEN* durch p_K gegeben. Es gelte: $h \leq t_K$.
Zeigen Sie: Die Anzahl der Blockzugriffe bei Ausführung eines Nested-Loop-Joins mit *KUNDEN* in der äußeren Schleife und dem Zugriff auf den Join-Index in der inneren Schleife kann niemals niedriger sein als bei Ausführung des Joins durch Scan des Join-Index.

Aufgabe 2: Algebraische Transformation (8 Punkte)

Gegeben seien die Relationen $R_1(\underline{A},B,C)$, $R_2(\underline{A},B,C)$ und $S(\underline{D},F)$ sowie die View $R(\underline{A},B,C)=R_1 \cup R_2$. Die unterstrichenen Attribute bilden den Primärschlüssel. Es gelten die folgenden Integritätbedingungen:

I1: $R(\underline{A},B,C)= R_1 \cup R_2$

I2: $\sigma_{[A>10]}(R_1)= R_1$

I3: $\sigma_{[A\leq 10]}(R_2)= R_2$

Die Relationen R_1 und R_2 bilden eine sogenannte “horizontale” Partitionierung der View *R*, die z.B. einer Einteilung in “billige” und “teure” Produkte entsprechen könnte.

Das Attribut *F* sei Fremdschlüssel in der Relation *S* auf die Relation R_2, d.h. es gelte:

I4: $\pi_{[A]}(R_2) \supseteq \pi_{[F]}(S)$

Gegeben sei die folgende Query:

$$\pi_{[F]}(\sigma_{[A\leq 10]} (R \,|x|_{[A=F]}\, S))$$

Transformieren Sie die Query so, daß unnötige Operationen eliminiert und essentielle Operationen auf möglichst kleinen Operanden ausgeführt werden. Berücksichtigen Sie neben gegebenen und ableitbaren Integritätsbedingungen die folgenden allgemeinen Regeln:

R1: $\pi_{[R.A]} (R \,|x|_{[R.A=S.B]}\, S) = \pi_{[R.A]} (R)$, wenn $\pi_{[R.A]} (R) \subseteq \pi_{[S.B]} (S)$

R2: $\sigma_{[P1]} (R \,|x|_{[P2]}\, S)= (\sigma_{[P1]} (R)) \,|x|_{[P2]}\, S$, wenn P1 nur Attribute aus R enthält
$\sigma_{[P1]} (R \,|x|_{[P2]}\, S)= R \,|x|_{[P2]}\, (\sigma_{[P1]} (S))$, wenn P1 nur Attribute aus S enthält

R3: $\sigma_{[P]} (R_1 \cup R_2)= \sigma_{[P]} (R_1) \cup \sigma_{[P]} (R_2)$

R4: $(R_1 \cup R_2) \,|x|_{[P]}\, S = (R_1 |x|_{[P]}\, S) \cup (R_2 |x|_{[P]}\, S)$

R5: $R \,|x|_{[P]}\, S = S \,|x|_{[P]}\, R$

Wichtig: Geben Sie bei jedem Transformationsschritt an, welche der Regeln Sie angewendet haben.