

Datenbanksysteme I

10. Übungsblatt, Sommersemester 2005

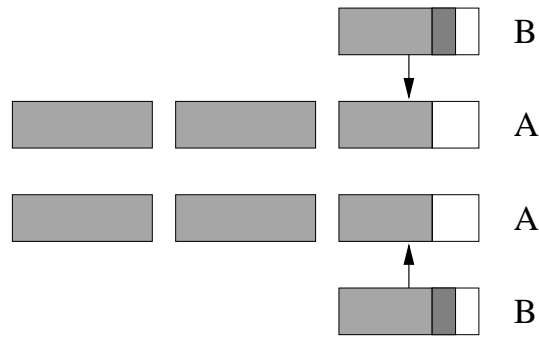
Besprechung am Mittwoch, den 29.06.2005

- Die Relation R enthält 6000 Tupel. Auf dieser Relation sollen nacheinander die Selektionen $\sigma_1, \sigma_2, \sigma_3$ und σ_4 ausgeführt werden. Da der Selektionsoperator kommutativ und assoziativ ist, spielt die Reihenfolge für das Ergebnis keine Rolle. Die Kosten (in Millisekunden) können für verschiedene Reihenfolgen aber durchaus verschieden sein. Tabelle 1 gibt die Selektivitäten und Auswertungskosten pro Tupel für jede Selektion an. Geben Sie die günstigste Auswertungsreihenfolge samt Kosten für die Gesamtauswertung an.

Selektion	Selektivität	Kosten pro Tupel
σ_1	$\frac{1}{2}$	5 ms
σ_2	$\frac{1}{20}$	1 ms
σ_3	$\frac{1}{15}$	2 ms
σ_4	$\frac{1}{10}$	5 ms

Tabelle 1: Selektivität und Kosten für Selektionen

- Für was steht das sogenannte ACID-Paradigma?
- Demonstrieren Sie anhand eines Beispiels, daß man die Strategien *force* und \neg *steal* nicht kombinieren kann, wenn parallele Transaktionen gleichzeitig Änderungen an Datenobjekten innerhalb einer Seite durchführen.
- Logeinträge werden aus Sicherheitsgründen meistens zweifach auf verschiedene Speichermedien (z.B. Platten) ausgeschrieben. Nehmen Sie an, daß die letzte Seite mit Logeinträgen noch nicht vollständig gefüllt ist, aber schon wichtige Logeinträge enthält. Der nächste Logeintrag wird an die Einträge dieser Seite hinten angehängt und die Seite wird auf beide Speichermedien gleichzeitig ausgeschrieben (was zum Überschreiben der alten Seite führt). Wenn in diesem Augenblick ein Fehler auftritt, können auf beiden Speichermedien die Einträge der letzten noch nicht vollständig aufgefüllten Seiten zerstört werden!



Die neuen Versionen (B) werden nicht korrekt geschrieben, die alten Versionen (A) sind aber überschrieben. Geben Sie ein Verfahren an, das diesen Fehler möglichst effizient behebt.

5. (Elektronische Abgabe)

(a) (4 Punkte)

Wieviele verschiedene Joinreihenfolgen gibt es für vier Relationen, wenn man sich nicht auf Left-Deep Trees beschränkt?

(b) (3 Punkte)

Die Relationen R_1 , R_2 und R_3 sollen gejoint werden. Dabei wird folgendes (vereinfachtes) Kostenmodell angenommen: $\text{Kosten}(R \bowtie S) = |R| \cdot |S|$ (wobei $|R|$ die Kardinalität (Größe) der Relation R ist). Die Kardinalität der Ergebnisrelation wird mit Hilfe von Selektivitäten bestimmt: $|R \bowtie S| = |R| \cdot |S| \cdot \text{sel}_{RS}$. Gegeben folgende Kardinalitäten und Selektivitäten, was sind die Kosten für die Auswertung des Plans $(R_2 \bowtie R_3) \bowtie R_1$? Kardinalitäten: $|R_1| = 100$, $|R_2| = 2000$, $|R_3| = 500$. Selektivitäten: $\text{sel}_{12} = \frac{1}{200}$, $\text{sel}_{13} = \frac{1}{50}$, $\text{sel}_{23} = \frac{1}{250}$.

(c) (1 Punkt)

Gibt es eine günstigere Auswertungsreihenfolge für die Relationen R_1 , R_2 und R_3 aus Teilaufgabe (b)?