

Datenbanksysteme II
SS 2012

Übungsblatt 8: Join-Verfahren

Besprechung: 02.-04.07.2012

Aufgabe 8-1 *I/O-Kosten: NBL-Join*

Gegeben seien zwei Relationen R und S , die jeweils eine Größe von 10.000 Blöcken besitzen. Im folgenden soll der Join $R \bowtie S$ mittels eines Nested-Block-Loop-Joins berechnet werden. Dabei wird als Cachestrategie Variante 3 (Siehe Skript) verwendet.

- (a) Berechnen Sie die benötigte Anzahl an Plattenzugriffen bei einer Cachegröße von 1.000 Blöcken.
- (b) Berechnen Sie die benötigte Cachegröße in Blöcken, um das Joinergebnis mit höchstens 100.000 Plattenzugriffen zu berechnen.
- (c) Berechnen Sie die benötigte Cachegröße in Blöcken, um das Joinergebnis mit höchstens 20.000 Plattenzugriffen zu berechnen.

Aufgabe 8-2 *Implementierung der Join-Operatoren: CPU-Kosten*

Es soll nun der Equi-Join der im Folgenden abgebildeten Relationen R und S berechnet werden.

R	S
1	1
2	2
3	3
6	4
7	5
9	6
10	8
11	10
12	13
13	14

- (a) Führen Sie den Join mittels des Nested-Block-Loop-Join durch. Wieviele Paare von Tupeln müssen dabei auf Erfüllung des Joinprädikates untersucht werden?
- (b) Führen Sie den Join mittels des Sort-Merge-Join durch. Wieviele Paare von Tupeln müssen dabei auf Erfüllung des Joinprädikates untersucht werden?

- (c) Führen Sie den Join mittels des einfachen Hash-Join mit Hashfunktion $h(x) = x \bmod 5$ durch. Wieviele Paare von Tupeln müssen dabei auf Erfüllung des Joinprädikates untersucht werden?
- (d) Führen Sie den Join mittels des Hash-Partitioned-Join (GRACE) durch. Dabei wird als Hashfunktion für die Blockpartitionierung die Funktion $h_B(x) = x \bmod 3$ verwendet. Auf den einzelnen Blöcken soll der Join mittels eines einfachen Hash-Joins mit der Hashfunktion $h(x) = x \bmod 2$ durchgeführt werden. Wieviele Paare von Tupeln müssen dabei auf Erfüllung des Joinprädikates untersucht werden?