

Übungen zur Vorlesung
Verteilte und Föderierte Datenbanksysteme

Tutoraufgaben (Besprechungstermin 23. 05. 2002, 14:15 Uhr)

T1 Gegeben seien die Relationen

ARBEITER(Name, Adresse, Gehalt, WerkNr)
MASCHINE(MaschNr, Type, WerkNr)

A) Die Relation ARBEITER soll nach WerkNr horizontal partitioniert werden. Jede der Partitionen soll zwei Kopien haben, einmal an dem Knoten München und einmal lokal am Werk selbst. Geben Sie eine gute Strategie zur Bearbeitung folgender Anfragen, die aus Frankfurt gestellt werden:

- Ermittle alle Mitarbeiter, die an dem Knoten Hamburg
- Das Durchschnittsgehalt aller Mitarbeiter
- Höchstbezahltester Mitarbeiter in folgenden Werken (Ulm, Stuttgart, Frankfurt, Berlin)
- Niedrigstes Gehalt in der Unternehmen.

T2 (Dadam 6-9) Es soll der Verbund zweier Relationen $R(A, B, C)$ und $S(A, D, E)$, die an den Knoten K_R bzw. K_S gespeichert sind, berechnet werden. Das Resultat ist an K_R verfügbar zu machen. Relation R umfasst 20.000 Tupel mit einer Tupelgröße von 200 Bytes und Relation S 60.000 Tupel mit einer Tupelgröße von 100 Bytes. Attribut A sei 10 Bytes groß. Es wird erwartet, dass ca. 1/3 der Tupel von S das Verbund-Prädikat erfüllt.

Angenommen, die eingesetzte Hashfunktion beim Hashfilter-Verbund würde bei einem Bitvektor von 200.000 Bit Länge 1% Pseudo-Treffer erzeugen und diese Rate würde sich bei einer Halbierung der Bitvektorgöße etwa jeweils vervierfachen. Wie groß (in Bits) muss man den Bitvektor mindestens wählen, damit der Hashfilter-Verbund günstiger als der Semi-Verbund ist?

Hausaufgaben (Abgabetermin 23. 05. 2002, 14:00 Uhr)

H1

Gegeben sei folgende vertikale Partitionierung:

$$R_1 := \pi_{A_1, \dots, A_j}(R)$$

$$R_2 := \pi_{A_1, A_{j+1}, \dots, A_n}(R)$$

der Relation $R(A_1, \dots, A_j, \dots, A_n)$.

- Führen Sie die Anfragetransformation für folgende Anfrage durch:

$$Q = \pi_{\{A_1, \dots, A_{j-2}, A_n\}} (\sigma_{R_1.A_1 = R_2.A_1} (R_1 \times R_2))$$

- Bestimmen Sie den Operatorbaum und führen Sie darauf Vereinfachungen und Restrukturierungen zur algebraischen Optimierung durch.

Zu beachten ist, dass die Relation R an einem Knoten gespeichert ist.

H2 Gegeben seien die Relationen **R** und **S** und ihre Ausprägungen:

| R | A | B |
|---|---|---|
| | 3 | 7 |
| | 1 | 1 |
| | 4 | 6 |
| | 7 | 7 |
| | 4 | 5 |
| | 6 | 2 |
| | 5 | 7 |

| S | B | C | D |
|---|---|---|---|
| | 9 | 8 | 8 |
| | 1 | 5 | 1 |
| | 9 | 4 | 2 |
| | 4 | 3 | 3 |
| | 4 | 2 | 6 |
| | 5 | 7 | 8 |

Die Relation R sei am Knoten **K_R** und S am Knoten **K_S** gespeichert.
Zu berechnen sei der **Verbund** zwischen R und S!

- Formulieren Sie in relationaler Algebra alle sinnvollen Möglichkeiten (Strategien) zu Berechnung des Verbundes Zwischen R und S mit Hilfe der Semi-Verbund Verfahren und setzen Sie diese jeweils in Form eines Semi-Verbund Algorithmus um und führen Sie diese Algorithmen an R und S aus.
- Der Nachrichtenumfang jeder Strategie ist die Anzahl aller zu übertragenden Attributwerte (alle Attribute sind hier vom gleichen Datentyp, so dass alle Werte den gleichen Umfang in einer Nachricht haben, d.h. die Attributgröße kann vernachlässigt werden).
Geben Sie für jede der zuvor betrachteten Strategien an, welchen Nachrichtenumfang sie bei der Berechnung des Verbunds zwischen R und S hat.

H3

- Der Hash-Filter-Verbund soll auf die in H2 gegebenen Relationen angewendet werden. Eine mögliche Hash-Funktion ist:
 $h(a) = (a \text{ MOD } n) + 1$, wobei **n = 5** die Länge des Bitvektors darstellt.
- An welchen Knoten soll die Berechnung durchgeführt werden, damit der Hashfilter-Verbund günstiger (hinsichtlich des Nachrichtenumfangs) als der Semiverbund ist.

Abgabe der Hausaufgaben deutlich mit Namen und VFDBS beschriftet, in den dafür vorgesehenen Übungskästen (beschriftet mit Übungen zu VFDBS) in den Garderoberraum am Eingang.

Bis zu drei Studentinnen/Studenten dürfen gemeinsam abgeben.