

Mobile und Verteilte Datenbanken

WS 2011/2012

Blatt 5

Lösung

Aufgabe 1: Abgeleitete horizontale Fragmentierung

Gegeben seien folgende Relationen:

ABT (ANR, ANAME, AORT, ABUDGET)
PERSONAL (PNR, PNAME, ANR, BERUF, GEHALT)
PROJEKT (PRONR, PRONAME, PROBUDGET)
PMITARBEIT (PNR, PRONR, DAUER).

Es handelt sich um das klassische Beispiel von Abteilungen, Projekten und ihren Mitarbeitern. Primärschlüssel sind unterstrichen; übereinstimmende Attributnamen kennzeichnen Fremdschlüssel-Primärschlüsselbeziehungen.

Abteilungen seien an drei Standorten vertreten: Kaiserslautern, Frankfurt und Leipzig.

Definieren Sie eine entsprechende horizontale Partitionierung von ABT sowie abgeleitete horizontale Partitionierungen für PERSONAL und PMITARBEIT. Welche Alternative besteht für PMITARBEIT zur Definition einer abgeleiteten horizontalen Fragmentierung? Wann wäre diese sinnvoll?

Lösung:

Horizontale Fragmentierung von ABT:

$$ABT_1 = \sigma_{AORT = "KL"}(ABT)$$

$$ABT_2 = \sigma_{AORT = "F"}(ABT)$$

$$ABT_3 = \sigma_{AORT = "L"}(ABT).$$

Abgeleitete horizontale Fragmentierung von PERSONAL (für den Fremdschlüssel ANR seien keine Nullwerte zulässig):

$$PERSONAL_1 = PERSONAL \bowtie ABT_1$$

$$PERSONAL_2 = PERSONAL \bowtie ABT_2$$

$$PERSONAL_3 = PERSONAL \bowtie ABT_3.$$

Abgeleitete horizontale Fragmentierung von PMITARBEIT:

$$PMITARBEIT_1 = PMITARBEIT \bowtie PERSONAL_1$$

$$PMITARBEIT_2 = PMITARBEIT \bowtie PERSONAL_2$$

$$PMITARBEIT_3 = PMITARBEIT \bowtie PERSONAL_3.$$

Im Falle einer horizontalen Fragmentierung von PROJEKT könnte für PMITARBEIT eine abgeleitete horizontale Fragmentierung auch bezüglich PROJEKT erfolgen. Dies wäre vorteilhafter als die auf PERSONAL ausgerichtete Fragmentierung, wenn häufiger eine Join-Berechnung mit PROJEKT als mit PERSONAL vorzunehmen ist.

Aufgabe 2: Hybride Fragmentierung

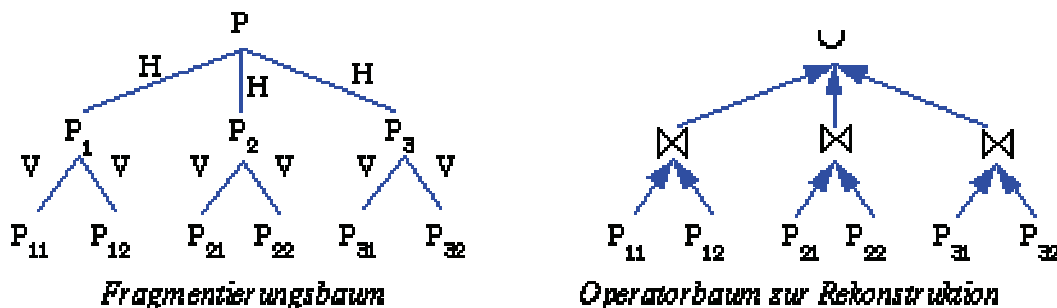
Auf der horizontalen Fragmentierung von PERSONAL aus der vorherigen Aufgabe soll zusätzlich eine vertikale Fragmentierung durchgeführt werden, so daß das GEHALT-Attribut in einem eigenen Fragment geführt wird. Geben Sie die Spezifikation einer solchen vertikalen Fragmentierung an. Zeichnen Sie den Fragmentierungsbaum sowie den Operatorbaum zur Rekonstruktion der PERSONAL-Relation.

Lösung:

Vertikale Fragmentierung von PERSONAL₁ (PERSONAL₂ und PERSONAL₃ analog):

$PERSONAL_{11} = \pi_{PNR, PNAME, ANR, BERUF}(PERSONAL_1)$

$PERSONAL_{12} = \pi_{PNR, GEHALT}(PERSONAL_1)$



Aufgabe 3: Daten-Lokalisierung (abgeleitete Fragmentierung)

Gegeben seien folgende Relationen:

PERSONAL (PNR, PNAME, BERUF, GEHALT)

PROJEKT (PRONR, PRONAME, PROBUDGET)

PMITARBEIT (PNR, PRONR, DAUER).

Weiterhin sei folgende horizontale Fragmentierung von Relation PROJEKT gegeben:

$PROJEKT_1 = \sigma_{PROBUDGET < 100000}(PROJEKT)$

$PROJEKT_2 = \sigma_{100000 \leq PROBUDGET \leq 800000}(PROJEKT)$

$PROJEKT_3 = \sigma_{PROBUDGET > 800000}(PROJEKT)$.

Für PMITARBEIT liege eine abhängige horizontale Fragmentierung vor:

$PMITARBEIT_1 = PMITARBEIT \bowtie PROJEKT_1$
 $PMITARBEIT_2 = PMITARBEIT \bowtie PROJEKT_2$
 $PMITARBEIT_3 = PMITARBEIT \bowtie PROJEKT_3$.

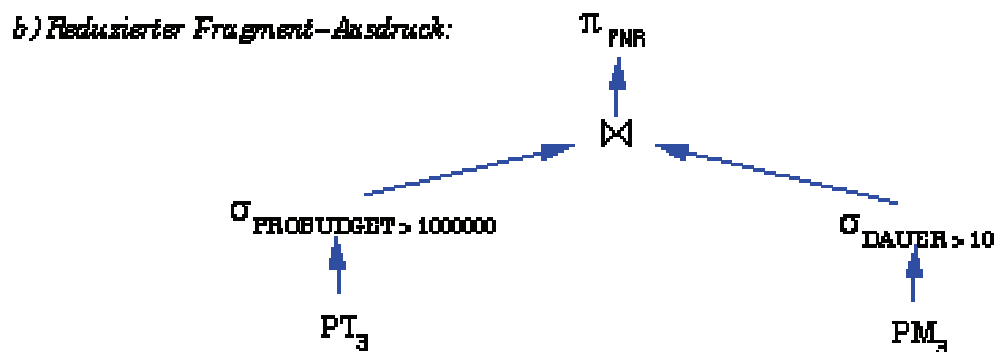
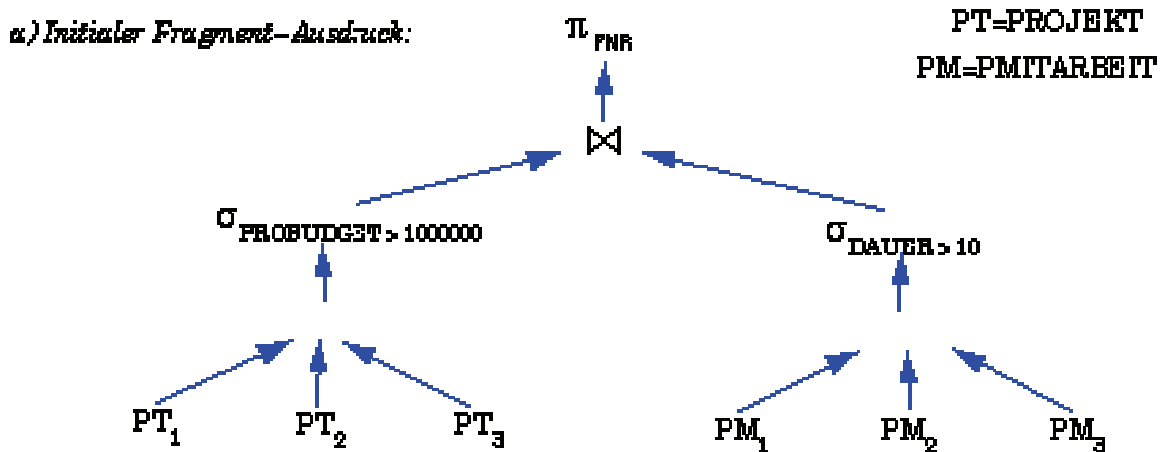
Bestimmen Sie für die Anfrage

```

SELECT PNR
FROM PROJEKT PT, PMITARBEIT PM
WHERE DAUER > 10 AND PROBUDGET > 1000000
AND PT.PRONR=PM.PRONR
  
```

zunächst den initialen Fragment-Ausdruck. Nehmen Sie algebraische Optimierungen zur weitestgehenden Reduzierung des Ausdrucks vor.

Lösung:



Aufgabe 4: Ship-Whole vs. Semi-Join vs. Bitvektor-Join

Auf den Relationen PERSONAL und PMITARBEIT (siehe vorherige Aufgabe) sei folgende Join-Query zu bearbeiten:

```

SELECT P.PNR, PNAME, BERUF, PRONR, DAUER
FROM PERSONAL P, PMITARBEIT PM
WHERE P.PNR=PM.PNR AND P.GEHALT > 60000
  
```

Es gelte $Card(\text{PERSONAL}) = 1000$, $Card(\text{PMITARBEIT}) = 1500$; beide Relationen seien an verschiedenen Knoten gespeichert. Die Anfrage soll an einem dritten Knoten

K initiiert werden; das Ergebnis ist dort auch auszugeben. Die Gehaltsbedingung soll von 20% der Angestellten erfüllt werden ($SF=0.2$); 25% der Angestellten sollen in keinem Projekt mitarbeiten.

Bestimmen Sie die Kommunikationskosten (#Nachrichten, #AW) für folgende Join-Strategien:

- Ship-Whole; Join-Berechnung an Knoten $K_{\text{PMITARBEIT}}$
- Ship-Whole; Join-Berechnung an Knoten K
- Semi-Join; Join-Bestimmung an Knoten K_{PERSONAL}
- Semi-Join; Join-Berechnung an Knoten K
- Bitvektor-Join; Join-Berechnung an Knoten K

Vor der Übertragung sollen alle anwendbaren Selektionen und Projektionen durchgeführt werden. Die Länge des Bitvektors soll 5 Attributwerten entsprechen; durch Anwendung des Bitvektors soll sich die zurückzuliefernde Tupelanzahl um 5% erhöhen.

Lösung:

Ship-Whole; Join-Berechnung an Knoten $K_{\text{PMITARBEIT}}$

- K -> $K_{\text{PMITARBEIT}}$: Nachricht zum Query-Start
- $K_{\text{PMITARBEIT}}$ -> K_{PERSONAL} : Anfordern der PERSONAL-Daten
- K_{PERSONAL} -> $K_{\text{PMITARBEIT}}$: $1000 \cdot 0.2 \cdot 3 = 600$ AW
- Join-Berechnung an $K_{\text{PMITARBEIT}}$.
- $K_{\text{PMITARBEIT}}$ -> K: Join-Ergebnis ($200 \cdot 0.75 \cdot 2 \cdot 5 = 1500$ AW)

=> #Nachrichten=4; #AW=2100.

Ship-Whole; Join-Berechnung an Knoten K

- K -> $K_{\text{PMITARBEIT}}$: Anfordern der PMITARBEIT-Daten
- K -> K_{PERSONAL} : Anfordern der PERSONAL-Daten
- $K_{\text{PMITARBEIT}}$ -> K: $1500 \cdot 3 = 4500$ AW
- K_{PERSONAL} -> K : $1000 \cdot 0.2 \cdot 3 = 600$ AW
- Join-Berechnung an K

=> #Nachrichten=4; #AW=5100.

Semi-Join; Join-Berechnung an Knoten K_{PERSONAL}

- K -> K_{PERSONAL} : Nachricht zum Query-Start
- K_{PERSONAL} -> $K_{\text{PMITARBEIT}}$: $1000 \cdot 0.2 = 200$ AW Verbundattributwerte
- $K_{\text{PMITARBEIT}}$ -> K_{PERSONAL} : $200 \cdot 0.75 \cdot 2 \cdot 3 = 900$ AW (Verbundpartner)
- Join-Berechnung an K_{PERSONAL} .
- K_{PERSONAL} -> K: Join-Ergebnis ($200 \cdot 0.75 \cdot 2 \cdot 5 = 1500$ AW)

=> #Nachrichten=4; #AW=2600.

Semi-Join; Join-Berechnung an Knoten K

- K -> $K_{\text{PMITARBEIT}}$: Anfordern der reduzierten Daten
- K -> K_{PERSONAL} : Anfordern der reduzierten Daten
- $K_{\text{PMITARBEIT}}$ -> K_{PERSONAL} : 750 AW Verbundattributwerte
- K_{PERSONAL} -> K : $750 \cdot 0.2 \cdot 3 = 450$ AW (Verbundpartner)
- K_{PERSONAL} -> $K_{\text{PMITARBEIT}}$: $1000 \cdot 0.2 = 200$ AW Verbundattributwerte
- $K_{\text{PMITARBEIT}}$ -> K : $200 \cdot 0.75 \cdot 2 \cdot 3 = 900$ AW (Verbundpartner)
- Join-Berechnung an K

=> #Nachrichten=6; #AW=2300.

Bitvektor-Join; Join-Berechnung an Knoten K

wie zuvor, jedoch wird in Schritten 3 und 5 jeweils der Bitvektor übertragen (-> 10 AW);

in Schritt 4 (6) erhöht sich der Übertragungsumfang auf ca. 475 (945) AW.

=> #Nachrichten=6; #AW=1430.

Beispiellösungen aus:

Rahm, E.: Mehrrechner-Datenbanksysteme, Addison-Wesley 1994.