

Aufgabenblatt 4

Optimierung und Benchmarking

- Abgabetermin: **Mittwoch, 27.06.07**
- Das Aufgabenblatt gilt als bestanden, wenn mindestens 20 der 40 Punkte erreicht werden.
- Die Aufgaben *sollen* in Zweiergruppen bearbeitet werden. Für Einzelabgaben ziehen wir 4 Punkte ab.
- Bitte verwendet für jede Aufgabe ein separates Blatt und beschriftet *jedes* Blatt der Abgabe mit Namen, Matrikelnummern und Seitenzahl!
- Abgabe: Auf Papier im Fach „Datenbanksysteme II“ im Foyer oder per E-Mail als pdf oder doc an dbs2@hpi.uni-potsdam.de mit cc an brigitte.hobro@hpi.uni-potsdam.de mit Subject „Abgabe DBS II: Aufgabenblatt n [Namen | Matrikelnummern]“
- **Die DB2-Dokumentation findest du unter:**
<http://publib.boulder.ibm.com/infocenter/db2luw/v9/index.jsp>

Aufgabe 1: Picasso Database Query Optimizer Visualizer

Verwende das Tool Picasso, um die Diagramme zu Ausführplänen und -kosten für die folgenden Anfragen zu betrachten. Verwende Plot Resolution 10 und betrachte die Pläne auf Operator Level.

a) Q1

```
select * from customer where c_custkey :varies
```

Hinweise:

- Das ist das Template für die Q1, Q2, Q3 aus Aufgabenblatt 3, Aufgabe 2.
 - Die Tabelle customer sollte jetzt wieder nach customer_pk organisiert sein. Statistiken nicht vergessen!
- 1) Wo waren also die Grenzen in Bezug auf c_custkey für die unterschiedlichen Ausführpläne? **2 P**
 - 2) Vergleiche das abgeschätzte und das reale Kostendiagramm (Comp Cost Diag und Exec Cost Diag). Was ist der Unterschied? **2 P**
 - 3) Vergleiche die Diagramme für Plot Resolution 10 und 30. Wie wirkt sich die Veränderung aus? **4 P**
- b) TPC-H Q3 (in Picasso: Load Query Template -> db2 -> q3.sql)

Hinweis:

- Die Tabelle customer sollte jetzt nach c_acctbal organisiert sein. Statistiken nicht vergessen!
- 1) Was sind die Unterschiede der Pläne? **6 P**
 - 2) Warum wird kein Merge Join verwendet? Was müsste man (vermutlich) ändern, damit ein Merge Join verwendet wird? **4 P**

c) TPC-H Q5 (in Picasso: Load Query Template -> db2 -> q5.sql)

Hinweis:

- Die Tabelle customer sollte jetzt nach c_acctbal organisiert sein. Statistiken nicht vergessen!
- 1) Was sind die Unterschiede der Pläne
 - an der Grenze 85 - 95 % Selektivität auf c_acctbal? 2 P
 - zwischen den Plänen für 20 und 40 % Selektivität auf c_acctbal? 2 P
 - 2) Vergleiche den Plan für 20 % Selektivität auf c_acctbal (auf Optimizer-Level 5) mit dem Plan auf Optimizer-Level 9. Was ist der Unterschied? 4 P

Hinweis: Im Picasso Jargon sucht ihr einen Foreign Plan Tree.

Hinweise zu Picasso:

- Diese Aufgabe *muss* auf einem Linux-Rechner bearbeitet werden.
- Dokumentation zu Picasso unter /local/picasso/Picasso1.0/PicassoDoc/index.htm
- Pfad zur Java-Installation: /local/jdk1.5.0_11WithJava3D/bin; also auf der Konsole (bash):
 PATH=/local/jdk1.5.0_11WithJava3D/bin:\$PATH
- Zum Aufruf des Clients: cd /local/picasso/Picasso1.0/PicassoRun/Unix; ./runClient.sh
- Anlegen einer DBConnection: DBConnection -> New
- Port eurer Instanz := 50050+ Instanz-Nummer
- Einige Interaktionen mit Picasso sind nur per Tasten- und Mauskombinationen verfügbar:
 Help -> Mouse-Key Mappings
- Query Template Descriptor: Kann man sich selbst ausdenken, sollte eindeutig sein, *muss* angegeben sein.

Aufgabe 2: Join-Kardinalität

Schätze die Kardinalität des Joins über $R(A, B) \bowtie S(B, C)$ ab unter Verwendung der Histogramme für $R(B)$ und $S(B)$, die die Häufigkeit der 4 häufigsten Werte angeben. Beide Relationen enthalten 20 unterschiedliche Werte im Attribut B ($V(R, B) = V(S, B) = 20$). 4 P

	0	1	2	3	4	andere
<i>R.B</i>	5	6	4	5		32
<i>S.B</i>	10	8	5		7	32

Aufgabe 3: Join-Reihenfolge

Gegeben seien die folgenden Relationen und deren Statistiken:

$W(A, B)$	$X(B, C)$	$Y(C, D)$	$Z(D, A)$
$T(W) = 100$	$T(X) = 200$	$T(Y) = 300$	$T(Z) = 400$
$V(W, A) = 20$			$V(Z, A) = 100$
$V(W, B) = 60$	$V(X, B) = 50$		
	$V(X, C) = 100$	$V(Y, C) = 50$	
		$V(Y, D) = 50$	$V(Z, D) = 40$

Bestimme die Join-Reihenfolge als Left Deep Tree. Gib dazu alle Tabellen des Algorithmus der Dynamischen Programmierung an. Was ist die optimale Join-Reihenfolge? Was sind deren Kosten? 10 P