

Aufgabenblatt 5 Anfrageoptimierung

- Abgabetermin: **Dienstag, 28.01.14 (23:59 Uhr)**
- Zur Prüfungszulassung muss ein Aufgabenblatt mit mind. 25% der Punkte bewertet werden und alle weiteren Aufgabenblätter mit mindestens 50% der Punkte.
- Die Aufgaben sollen in Zweiergruppen bearbeitet werden.
- Abgabesystem unter
<http://www.dcl.hpi.uni-potsdam.de/submit>
 - ausschließlich pdf-Dateien
 - eine Datei pro Aufgabe namens Aufgabe-<aufgabenNr>.pdf
 - jedes Blatt beschriftet mit Namen

Aufgabe 1: Kardinalitätsschätzung

Gegeben seien die folgenden Relationen und deren Statistiken:

| | | | |
|----------------|-----------------|----------------|-----------------|
| $W(a, b)$ | $X(b, c)$ | $Y(c, d)$ | $Z(d, a)$ |
| $T(W) = 300$ | $T(X) = 600$ | $T(Y) = 900$ | $T(Z) = 1200$ |
| $V(W, a) = 30$ | | | $V(Z, a) = 300$ |
| $V(W, b) = 60$ | $V(X, b) = 50$ | | |
| | $V(X, c) = 100$ | $V(Y, c) = 50$ | |
| | | $V(Y, d) = 60$ | $V(Z, d) = 40$ |

Schätze die Kardinalität der Ergebnisrelationen der folgenden Ausdrücke:

9 P

- a) $W \bowtie X \bowtie Y \bowtie Z$
- b) $\sigma_{a=10}(W)$
- c) $\sigma_{c=20}(Y)$
- d) $\sigma_{c=20}(Y) \bowtie Z$
- e) $W \times Y$
- f) $\sigma_{d>10}(Z)$
- g) $\sigma_{a=1 \wedge d=2}(Z)$
- h) $\sigma_{c>1 \wedge d=2}(Y)$
- i) $X \bowtie_{X.b=Z.d} Z$

Aufgabe 2: Join-Kardinalität

Gegeben sind zwei Relationen $R(A, B)$ und $S(B, C)$. Beide Relationen enthalten 20 unterschiedliche Werte im Attribut B , wobei die Werte in R den Werten in S entsprechen. Es gilt also $V(R, B) = V(S, B) = 20$. Die Werteverteilungen in $R(B)$ und $S(B)$ sind durch folgendes Histogramm beschrieben, welches die Häufigkeit der 4 häufigsten Werte angibt:

| | | | | | | |
|-------|----|---|---|---|---|--------------|
| | 0 | 1 | 2 | 3 | 4 | andere Werte |
| $R.B$ | 5 | 6 | 4 | 5 | * | 32 |
| $S.B$ | 10 | 8 | 5 | * | 7 | 48 |

Die mit * gekennzeichneten Werte gehören nicht zu den vier jeweils häufigsten Werten, sondern zu den "anderen Werten". Schätze nun unter Verwendung des Histogramms die Kardinalität des Joins über $R(A, B) \bowtie S(B, C)$ ab. Vergleiche anschließend die berechnete Abschätzung mit der einfachen Abschätzung, die eine Gleichverteilung aller 20 Attributwerte annimmt. **5 P**

Aufgabe 3: Join-Reihenfolge

Gegeben seien die folgenden Relationen und deren Statistiken:

| $E(a, b, c)$ | $F(a, b, d)$ | $G(a, c, d)$ | $H(b, c, d)$ |
|------------------|-----------------|-----------------|-----------------|
| $T(E) = 1000$ | $T(F) = 2000$ | $T(G) = 3000$ | $T(H) = 4000$ |
| $V(E, a) = 1000$ | $V(F, a) = 50$ | $V(G, a) = 50$ | |
| $V(E, b) = 50$ | $V(F, b) = 100$ | | $V(H, b) = 40$ |
| $V(E, c) = 20$ | | $V(G, c) = 300$ | $V(H, c) = 100$ |
| | $V(F, d) = 200$ | $V(G, d) = 500$ | $V(H, d) = 400$ |

Bestimme die Join-Reihenfolge als Left Deep Tree. Gib dazu alle Tabellen des Algorithmus der Dynamischen Programmierung an. Was ist die optimale Join-Reihenfolge, wie hoch sind deren Kosten und welche Kardinalität hat der Join am Ende? **10 P**