

8. Übungsblatt

Ausgabe: 8. Januar 2014 · Besprechung: 15. Januar 2014

1 Paralleles Datenbanksystem

In einem parallelen Datenbanksystem sind zwei Relationen gespeichert:

$$\begin{aligned} & \text{Employees} (\underline{\text{eid}} : \text{integer}, \text{did} : \text{integer}, \text{salary} : \text{real}) \\ & \text{Departments} (\underline{\text{did}} : \text{integer}, \text{mgrid} : \text{integer}, \text{budget} : \text{integer}) \end{aligned}$$

Dabei ist *mgrid* ein Fremdschlüssel auf *Employees*. Beide Relationen haben eine Tupelgröße von 20 Bytes. Die Felder *salary* und *budget* enthalten gleichverteilte Daten zwischen 0 und 1.000.000. Die Relation *Employees* ist 100.000 Datenbankseiten groß; *Departments* besteht aus 5.000 Seiten (Seitengröße: 4.000 Bytes). Es gibt keine Indizes und alle (Equi-)Joins werden als *Sort-Merge-Join* ausgeführt. Beide Relationen sind anfangs *round-robin*-partitioniert. Das System verfügt über 10 Prozessoren, die jeweils über einen *Buffer Pool* von 100 Seiten.

Im folgenden wollen wir für diese Konfiguration die Kosten einer Anfrageauswertung abschätzen. Dabei sollen Kosten in Form von "Zeit" berechnet werden, wobei wir annehmen

- eine I/O-Operation kostet/dauert t_{IO} ,
- eine Seite über das Netzwerk zu bewegen kostet/dauert t_{net} .

Bestimmen Sie für die folgenden Anfragen

- (a) die aufgewendeten **Gesamtkosten** sowie
- (b) die **Ausführungszeit** der Anfrage (Prozessoren arbeiten teilweise parallel; dann ist die Ausführungszeit bestimmt durch den *langsamsten* Prozessor).

Betrachten Sie folgende Anfragen:

1. Gesucht ist der Mitarbeiter mit dem höchsten Gehalt.
2. Gesucht ist der Mitarbeiter mit dem höchsten Gehalt, der in einer Abteilung mit einem $\text{budget} < 150.000$ arbeitet.

2 Paralleles Datenbanksystem—*Range Partitioning*

Bestimmen Sie nochmals die Kosten für das Szenario von Aufgabe 1. Nehmen Sie dabei jedoch an, dass beide Tabellen *range-partitioned* bezüglich der Attribute *salary* bzw. *budget* sind.

3 Paralleles Datenbanksystem—Skalierbarkeit

Wie verändern sich die Kosten für das System aus Aufgabe 1, wenn es über 100 (statt 10) Prozessoren verfügt?