

## Übung zur Vorlesung „Informationssysteme: Verteilte und Web-Datenbanken“

Richard Kuntschke (richard.kuntschke@in.tum.de)

### Blatt 9

**Abgabe: Dienstag, 21. Dezember 2004 bis 12.00 Uhr** im Raum MI 02.11.018 im entsprechend beschrifteten Fach.

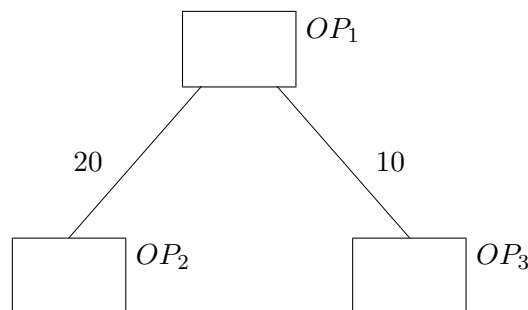
#### Aufgabe 1 - Generalized Hash Teams (8 Punkte)

In der Vorlesung wurde das Verfahren der Generalized Hash Teams für Join und Gruppierung vorgestellt. Ziel dieses Verfahrens ist es, mehrfaches Schreiben von Zwischenergebnissen, die aufgrund von wiederholt anzuwendenden Partitionierungen für Join- und Gruppierungsoperationen nötig wären, durch ein geschicktes Partitionierungsverfahren (indirektes Partitionieren) zu vermeiden. Hierdurch werden also I/O-Kosten eingespart. Überlegen Sie sich Möglichkeiten, dieses Verfahren im Bereich verteilter und paralleler Datenbanken anzuwenden. Für eine verteilte Datenbank ist die Reduktion von Kommunikationskosten das wichtigste Optimierungsziel. Parallele Datenbanken werden in lokalen Umgebungen mit hohen Übertragungsleistungen des eingesetzten Netzwerks (Rechnercluster oder Parallelrechner) betrieben, wobei das Hauptaugenmerk auf einer möglichst hohen Parallelität in der Anfrageauswertung liegt.

Lesen Sie sich zum tieferen Verständnis des obigen Verfahrens die entsprechende wissenschaftliche Veröffentlichung „Generalized Hash Teams for Join and Group-by“ durch (ein Link ist auf der WWW-Seite zu dieser Übung zu finden).

#### Aufgabe 2 - Antwortzeit-Kostenmodell (6 Punkte)

Berechnen Sie anhand des in der Vorlesung vorgestellten Modells die Antwortzeit für folgenden Plan:



Die Operationen  $op_1$ ,  $op_2$  und  $op_3$  haben folgende Ausführungsdauern: (60, 50, 100). Das erste Ergebnistupel liefern sie respektive nach (20, 30, 20). Die Netzwerkkosten sind in der Abbildung angegeben.

### **Aufgabe 3 - Vergleich: Kostenmodelle** (6 Punkte)

Herkömmliche Kostenmodelle berechnen den Gesamtaufwand, den eine Anfragebearbeitung nach sich zieht. Antwortzeitkostenmodelle interessiert im Gegensatz dazu nur die tatsächlich benötigte Zeit zur Ausführung der entsprechenden Anfrage. Bestimmen Sie eine Beispielanfrage für eine von Ihnen vorgegebene verteilte Datenbank zusammen mit einem dazugehörigen Plan, bei der Ihrer Meinung nach die Diskrepanz zwischen der Vorhersage eines gewöhnlichen Kostenmodells und eines Antwortzeitkostenmodells relativ hoch ist.

#### **Allgemeine Hinweise**

- Die Aufgaben der Übungsblätter sollen gemeinsam von den Mitgliedern eines Teams bearbeitet werden. Die Teameinteilung (je 3 oder 4 Teilnehmer) ist während des Semesters fest!
- Schreiben Sie bitte auf jedes Blatt die Namen aller Team-Mitglieder und die Team-Nummer. Jedes Team-Mitglied sollte die bearbeiteten Aufgaben des Teams „vorrechnen“ können (d. h. der Lösungsansatz des Teams und evtl. dabei auftretende Probleme sollten bekannt und verstanden sein).
- Die bearbeiteten Aufgaben können handschriftlich (bitte leserlich schreiben!) oder in getippter Form abgegeben werden. Bitte heften Sie alle Blätter einer Abgabe zusammen.
- Bitte beachten Sie, dass Abgaben nur dann korrigiert werden können, wenn Sie rechtzeitig zum Abgabetermin abgegeben wurden, auf der Abgabe die Namen aller Team-Mitglieder und die Team-Nummer angegeben sind und die Team-Einteilung eingehalten wurde.