

Mobile und Verteilte Datenbanken

WS 2010/2011

Blatt 5

Aufgabe 1: Parallele Join-Berechnung

Relation R (1 Million Tupel) sei an 4 Rechnern ($n=4$), Relation S (100.000 Tupel) an 2 Rechnern ($m=2$) gespeichert. Wie hoch ist der Kommunikationsumfang (in MB) für die Datenumverteilung zur parallelen Join-Berechnung

- für dynamische Replikation
- für dynamische Partitionierung und 5 Join-Rechner ($p=5$)
- für dynamische Partitionierung, wenn für S das Verteil- mit dem Join-Attribut übereinstimmt, nicht jedoch für R?

Die Tupelgröße betrage 100 B.

Aufgabe 2: TID-Hash-Join

Eine Reduzierung des Speicherbedarfs von Hash-Joins ergibt sich, wenn in der Hash-Tabelle anstelle der vollständigen Sätze nur die Schlüsselwerte für das Join-Attribut sowie die Verweise (tuple identifiers, TID) auf die Sätze gespeichert werden. Diskutieren Sie Vor- und Nachteile eines solchen Ansatzes. Welche Auswirkungen ergeben sich für die parallele Join-Bearbeitung in Shared-Nothing-Systemen ?

Aufgabe 3: Paralleler Hash-Join mit Überlaufbehandlung an Datenknoten

Geben Sie einen Algorithmus für einen parallelen Hash-Join in Shared-Nothing-Systemen an, bei dem die Überlaufbehandlung gemäß dem GRACE-Ansatz an den Datenknoten erfolgt.

Aufgabe 4: Paralleler Hash-Join bei Shared-Everything

Wie können die Algorithmen zur parallelen Hash-Join-Berechnung für Shared-Everything abgewandelt werden?