

VerteilteWebInf Hausaufgabe 7

Gruppe 6

November 25, 2014

Aufgabe 1

- a) Annahme: Es wären zwei Joinpartner $r \in R$ und $s \in S$ in unterschiedlichen Partitionen \mathcal{R}_i und \mathcal{R}_j .
Dann gilt $h(r.A) \bmod p \neq h(s.B) \bmod p$ bei gleicher Hashfunktion h , also gilt $r.A \neq s.B$, was ein Widerspruch dazu ist, dass r und s Joinpartner sind.
- b) unter der Annahme, dass jedem Knoten genau eine Partition zugeteilt wird muss Knoten i
 $\frac{p-1}{p} |\mathcal{R}_i| + \frac{p-1}{p} |\mathcal{S}_i|$ Tupel verschicken.
Zu addieren sind noch die Ergebnistupel aus dem Join von $\mathcal{R}_{j,i}$ und $\mathcal{R}_{j,i} \forall j$, die gegebenenfalls an eine andere Station gesendet werden müssen.
- c) Bei einem Join mit Broadcast verschickt jeder Knoten $|\mathcal{R}|$ Tupel.
Zu addieren sind noch die Ergebnistupel aus dem Join von \mathcal{R} und \mathcal{S}_i , die gegebenenfalls an eine andere Station gesendet werden müssen.
- d) $\mathcal{R} < \frac{p-1}{p} |\mathcal{R}_i| + \frac{p-1}{p} |\mathcal{S}_i|$
Annahme: R und S sind gleichmäßig auf die p Knoten verteilt (d.h. die Partitionen sind alle gleich groß)
 $\mathcal{R} < \frac{p-1}{p} \cdot \frac{1}{p} |\mathcal{R}| + \frac{p-1}{p} \cdot \frac{1}{p} |\mathcal{S}|$
 $\iff \mathcal{R} < \frac{p-1}{p^2 - p - 1} |\mathcal{S}|$
- e) Alle Knoten $i \neq k$ schicken gleichzeitig D Daten an Knoten k . Dieser empfängt mit Bandbreite b .
Die gesamte Zeit, bis Knoten k alle Daten vollständig erhalten hat, ist somit $t_{vj} = \frac{(n-1)D}{b}$.
- f) Pro Knoten müssen jetzt $p(n-1)$ Partitionen an Knoten k verschickt werden, also: $t = (n-1) \frac{(n-1)D}{b}$.
Es dauert also $(n-1)$ mal länger als beim verteilten Join.
- g) Knoten 1 muss jetzt doppelt so viele Daten senden, die Zeit insgesamt ergibt sich dann zu $t = (n-2) \frac{D}{b} + \frac{2D}{b}$
- h) Wenn $p > n$ muss jeder Knoten u.U. nicht alle seine Tupel senden, sondern es reicht, eine seiner Partitionen zu senden.

Aufgabe 2

- a) *min_{data}.lp*
- Partition 0: Knoten 0
 - Partition 1: Knoten 0
 - Partition 2: Knoten 3

- Partition 3: Knoten 0
- Partition 4: Knoten 1
- Partition 5: Knoten 0
- Partition 6: Knoten 0
- Partition 7: Knoten 0
- Partition 8: Knoten 0
- Partition 9: Knoten 0
- Partition 10: Knoten 0
- Partition 11: Knoten 0
- Partition 12: Knoten 0
- Partition 13: Knoten 0
- Partition 14: Knoten 0
- Partition 15: Knoten 0

b) $\min_{\text{antwortzeit}}.lp$

- Partition 0: Knoten 3
- Partition 1: Knoten 0
- Partition 2: Knoten 3
- Partition 3: Knoten 2
- Partition 4: Knoten 1
- Partition 5: Knoten 2
- Partition 6: Knoten 2
- Partition 7: Knoten 3
- Partition 8: Knoten 0
- Partition 9: Knoten 0
- Partition 10: Knoten 0
- Partition 11: Knoten 1
- Partition 12: Knoten 1
- Partition 13: Knoten 0
- Partition 14: Knoten 0
- Partition 15: Knoten 3

c) *Aufgabe a:*

gesamte Übertragungskosten: 313

Übertragungszeit: 264

Aufgabe b:

gesamte Übertragungskosten: 328

Übertragungszeit: 95

insgesamt sieht man also, dass bei Aufgabe a) die gesamten Übertragungskosten minimiert sind, während bei b) die Übertragungszeit minimiert wurde.

Besonders auffällig ist hierbei, dass bei nur geringer höheren Gesamtübertragungskosten die Übertragungszeit dramatisch sinkt, weshalb in diesem Fall vermutlich b) zu bevorzugen wäre.