

5. Übungsprojekt

Materialisierte Sichten

Ein wichtiger Aspekt beim Verwenden von materialisierten Sichten ist die Auswahl dieser aus der Menge aller möglichen Sichten, um den Trade-Off zwischen Speicherbedarf und durchschnittlicher Anfrageauswertungszeit auszugleichen. In der Vorlesung haben Sie dazu einen Verbund von Sichten kennengelernt, der Sichten entsprechend der verwendeten Group By Attribute anordnet, so dass eine Sicht jeweils aus jeder Untergeordneten berechnet werden kann. Aus diesem Verbund muss nur das unterste Element (die fein granularste Sicht) gespeichert werden, um alle Anfragen beantworten zu können. Das Speichern jedes weiteren Elements verringert die durchschnittliche Anfrageauswertungszeit.

In dem Artikel [1] wird ein Greedy Algorithmus vorgestellt, der eine Auswahl von k Sichten trifft, um die durchschnittliche Anfragezeit zu minimieren. Es wird dabei vereinfachend angenommen, dass die Zeit zur Auswertung einer Anfrage auf einer Sicht proportional zu Größe der Sicht ist. Ihr Übungsprojekt besteht aus den folgenden Teilaufgaben:

1. Verstehen Sie den Greedy Algorithmus zur Sichtenauswahl in [1, Abschnitt 4.1].
2. Definieren Sie einen Verbund von Sichten für das AWC Data Warehouse entsprechend [1, Abschnitt 2], der mindestens acht Knoten enthält. Bestimmen Sie die Kosten für jede Sicht in dem Verbund als Anzahl der enthaltenen Zeilen für das AWC Data Warehouse.
3. Führen Sie den Greedy Algorithmus für diesen Verbund durch wie in [1, Beispiel 4.1].

Literatur

- [1] Venky Harinarayan, Anand Rajaraman und Jeffrey D. Ullman: Implementing Data Cubes Efficiently. Proceedings of the 1996 ACM SIGMOD International Conference on Management of Data. ACM Press, 1996, Seiten 205-216.