

4. Übungsblatt

Ausgabe: 47. Kalenderwoche · Besprechung: 48/49. Kalenderwoche

Aufgabe 0

- Was ist ein Index in einem DWH?
- Was ist ein Join in einem DWH?
- Welche Indexstrukturen kennen Sie?
- Was ist ein Index Nested Loop?

Aufgabe 1

Gegeben sei das SSB Schema aus dem angehängten Paper [OOCR09]. Ferner sei folgende Anfrage gegeben:

```
select count(*), P_CATEGORY
from LINEORDER, PART, CUSTOMER, DATE, SUPPLIER
where D_YEAR between 1992 and 1995 and C_REGION="EUROPE" and S_REGION="AMERICA"
and LO_CUSTKEY=C_CUSTKEY and LO_PARTKEY=P_PARTKEY and LO_ORDERDATE=D_DATE and
LO_SUPPKEY=D_SUPPKEY
group by P_CATEGORY
```

In der Vorlesung haben Sie unterschiedliche Strategien zur Anfrageausführung kennengelernt. Erklären Sie die Strategien an der gegebenen Anfrage! Erläutern Sie welche Indexstrukturen bei den unterschiedlichen Strategien angewendet werden! Beschreiben Sie wie die Joins jeweils ausgeführt werden! Nennen Sie die Vor- und Nachteile der jeweiligen Strategien!

Aufgabe 2

Wir wollen einen Join in einem DWH System simulieren. Dafür sei wieder das Schema und die Anfrage aus Aufgabe 1 gegeben. Ferner sind die Tabellen in der gepackten Datei im Anhang

gegeben. Implementieren Sie in Java einen statischen Hashjoin! Das heißt Sie haben nur die gegebene Anfrage und die Tabellen. Implementieren Sie sowohl Strategie 1 als auch Strategie 2 aus der Vorlesung! Lesen Sie die Tabellen in Arrays ein (Die Spalten sind durch ein | getrennt)! Für Strategie 1 implementieren Sie Hash Indices auf die Fremdschlüsseln in der Faktentabelle. Für den Join durchlaufen Sie die Dimensionstabellen sequentiell. Für die Strategy 2 implementieren Sie Hash Indices (Hash Tabellen) auf den Primärschlüsseln der Dimensionstabellen. Beim Join durchlaufen Sie die Faktentabelle sequentiell. Hinweis: Für Strategie 1 werden normalerweise noch Indices auf den Wertespalten der Dimensionstabellen angelegt. Darauf verzichten wir. Dadurch müssen Sie die Dimensionstabellen sequentiell durchlaufen, um die passenden Tupel zu finden. Vergleichen Sie die Laufzeit der jeweiligen Strategien (ohne Einlesen und Indizieren den Daten)!

Literatur

- [OOCR09] Patrick E. O’Neil, Elizabeth J. O’Neil, Xuedong Chen, and Stephen Revilak. The star schema benchmark and augmented fact table indexing. In Raghunath Othayoth Nambiar and Meikel Poess, editors, *TPCTC*, volume 5895 of *Lecture Notes in Computer Science*, pages 237–252. Springer, 2009.