

Aufgabenblatt 3

Indexstrukturen und Anfrageausführung

- Abgabetermin: **Mittwoch, 13.06.07**
- Das Aufgabenblatt gilt als bestanden, wenn mindestens 25 der 50 Punkte erreicht werden.
- Die Aufgaben *sollen* in Zweiergruppen bearbeitet werden. Für Einzelabgaben ziehen wir 4 Punkte ab.
- Bitte verwendet für jede Aufgabe ein separates Blatt und beschriftet *jedes* Blatt der Abgabe mit Namen, Matrikelnummern und Seitenzahl!
- Abgabe: Auf Papier im Fach „Datenbanksysteme II“ im Foyer oder per E-Mail als pdf oder doc an dbs2@hpi.uni-potsdam.de mit cc an brigitte.hobro@hpi.uni-potsdam.de mit Subject „Abgabe DBS II: Aufgabenblatt n [Namen | Matrikelnummern]“
- **Die DB2-Dokumentation findest du unter:**
<http://publib.boulder.ibm.com/infocenter/db2luw/v9/index.jsp>

Aufgabe 1: Erzeugen der Datenbank und Laden der Daten

- Erzeuge in deiner Instanz eine Datenbank unter Verwendung der Anweisungen in `create_tpch_db.sql` (unter `lehrveranstaltungen\DBSII_naumann`). Die Tablespace werden auf `/data1` und `/data2` angelegt. Ersetze `dbs2in1` bzw. `dataDbs2in1` durch deine Instanz.
- Erzeuge das TPC-H Schema mit dem Skript `create_tpch_schema.sql`.
- Lade die Daten, setze die Integrität und sammle die Statistiken unter Verwendung der Anweisungen in `load_tpch_data.sql`. Hinweis: Miss die Zeit zum Laden der Daten in die `customer`-Tabelle (benötigt in Aufgabe 3).

Die Daten liegen auf `isis` im Verzeichnis `/data2/tpchData`.

Aufgabe 2: Indexe und Workloads

Betrachte die Relation `Customer` und ihre Indexe. Erstelle einen weiteren, sortierten Index auf dem Attribut `c_acctbal` und sammle alle Statistikinformationen.

Der Workload sei wie folgt gegeben:

Q1: `select * from customer where c_custkey < 50000`

Q2: `select * from customer where c_custkey < 100000`

Q3: `select * from customer where c_custkey < 140000`

Q4: `select * from customer order by c_acctbal`

Q5: `select * from customer where c_acctbal between 0 and 1000 order by c_acctbal`

- a) Betrachte die Ausführpläne zu Q1, Q2 und Q3.
- 1) Gibt es einen echten Unterschied der Ausführpläne zu Q1 und Q2? Warum (nicht)? **4 P**
 - 2) Was ist der Unterschied zwischen den Ausführplänen zu Q2 und Q3? Warum werden deiner Meinung nach die Pläne so gewählt? **4 P**
- b) Betrachte die Ausführpläne zu Q4 und Q5. Was ist der Unterschied? Warum werden deiner Meinung nach die Pläne so gewählt? **4 P**
- c) Reorganisiere die Tabelle customer nach dem erstellten Index über c_acctbal:
reorg table <schema>.customer index <schema>.<indexName>
- 1) Welche Veränderungen erwartest du (wenn überhaupt)? **2 P**
 - 2) Betrachte die Ausführpläne zu allen fünf Anfragen. Welche Veränderungen gibt es? War die Reorganisation eine gute Wahl für den gegebenen Workload? **6 P**

Aufgabe 3: Import vs. Load

Es gibt zwei Varianten Daten in eine DB2-Datenbank zu bringen – Import und Load. Per Load haben wir bereits alle Daten in das Schema geladen.

Lege nun eine zweite, leere customer-Tabelle an und *importiere* die Daten in diese Tabelle.

- a) Vergleiche die Laufzeiten für beide Varianten. **4 P**
- b) Wie erklärst du dir diesen Unterschied? Nutze die DB2-Doku für deine Antwort. **4 P**

Aufgabe 4: Mehrattributige Indexe I (Verwendung von Zusatzattributen)

Betrachte folgende Anfragen und deren Ausführpläne mit folgenden Indexten:

Anfragen

Q1: **select** o_orderkey
from dbs2in1.orders **where** o_orderkey > 100000

Q2: **select** o_orderkey, o_custkey
from dbs2in1.orders **where** o_orderkey > 100000;

Q3: **select** o_orderkey, o_custkey, o_orderstatus
from dbs2in1.orders **where** o_orderkey > 100000;

Q4: **select** o_orderkey, o_custkey, o_orderstatus, o_totalprice
from dbs2in1.orders **where** o_orderkey > 100000;

Indexe

I1: **create unique index** orders_idx1 **on** orders (o_orderkey) include (o_custkey)
collect detailed statistics

I2: **create unique index** orders_idx2 **on** orders (o_orderkey) include (o_custkey, o_orderstatus)
collect detailed statistics

- a) Welche Ausführpläne werden für die Anfragen gewählt
- 1) mit Index I1
 - 2) mit Index I1 und I2.
- Erkläre die Wahl der Indexe. 6 P
- b) Wird ein Index verwendet für Q4? Wie müsste man die Anfrage verändern, damit ein Index verwendet wird? (2 Möglichkeiten) Begründe deine Antwort. 4 P

Aufgabe 5: Mehrattributige Indexe II

- a) Lege den folgenden Index an:
- ```
create index orders_idx3 on orders (o_orderkey, o_custkey) collect detailed statistics
```
- Was passiert? Was sagt dir das über Indexe mit Zusatzattributen in DB2? 4 P
- b) Betrachte die Ausführpläne für die folgenden Anfragen.
- Q5: **select** o\_orderkey, o\_custkey **from** dbs2in1.orders **where** o\_orderkey > 100000
- Q6: **select** o\_orderkey, o\_custkey **from** dbs2in1.orders **where** o\_custkey > 100000
- Werden Indexe verwendet? Was sagt dir das über multidimensionale Indexe in DB2? 4 P
- c) Lege den folgenden Index an:
- ```
create index orders_idx4 on orders (o_custkey, o_orderkey) collect detailed statistics
```
- Verändert sich der Ausführplan für Q6? Was sind die Unterschiede? Was sagt dir das über multidimensionale Indexe in DB2? 4 P