



3. Pflichtübung

Einführung in die Datenbanksysteme Datenbanken für die Bioinformatik

Annika Hinze, Manuel Scholz

Bemerkungen

- Dies ist die dritte der drei Pflichtübungen im Rahmen der Vorlesung. Für den Scheinerwerb ist ein Mittelwert von mehr als 60% notwendig.
- Die Übung muss bis spätestens Freitag den 7.7.06 um 12.00 Uhr im Fach des jeweiligen Tutors (bzw. im Büro) abgegeben werden.
- Bitte alle Namen der beteiligten Personen der Gruppe auf die Lösungszettel schreiben. Die Abgabe erfolgt in den Projektgruppen.
- Das Kopieren von Lösungen wird bei allen beteiligten Gruppen als Täuschungsversuch gewertet.

Aufgabe (42 Punkte)

Gegeben sei ein Datenbanksystem das die Variante der B+ Bäume verwendet in der die Blätter des Baumes Daten enthalten und nicht nur Pointer zu den Dateneinträgen (Oracle: index organized table). Der Index ist für ein eindeutiges (unique) Attribut *b* angelegt, welches keine Duplikate enthält. Das Datenbanksystem hat die folgenden Eigenschaften:

- Blockgröße 4096 Bytes. Header: 96 Bytes. PCTFree:15%.
 - Jeder Eintrag ist 400 Bytes groß.
 - Pointer: 12 Bytes. Row-id (=Datenpointer): 12 Bytes. Search key: 8 Bytes.
 - Die Tabelle hat 10.000.000 Einträge.
- Wie groß wird der Index maximal? Geben Sie die Antwort als Anzahl der Blöcke und als Byte Anzahl. Zeigen Sie jeden Schritt in der Berechnung. (13 Punkte)
 - Wie groß wird der Index maximal wenn in jedem Blattknoten nur Pointer anstatt der Datengespeichert werden? Geben Sie die Antwort als Anzahl der Blöcke und als Byte Anzahl. Zeigen Sie jeden Schritt in der Berechnung.(9 Punkte)
 - Betrachten Sie eine Bereichsanfrage die auf 15% der Daten zugreift. Die Daten sind sequentiell nach dem search key geordnet. Ein Beispiel ist

```
select * from <table name> where b between x and y.
```

 Welche der folgenden Zugriffsformen wird am schnellsten sein? Untersuchen Sie für jede der Formen die maximale Anzahl der Blockzugriffe. Eventuelle Annahmen müssen explizit angegeben werden. (16 Punkte):
 - Zugriff über B+ Baum mit Datenblättern (wie in a).
 - Zugriff über B+ Baum (wie in b).
 - Zugriff über Hash auf Attribut *b* (extensible hash, directory in Hauptspeicher).
 - Zugriff (ohne Index) durch Lesen des gesamten Heaps.
 - Was würde sich verändern in (c) wenn ein B* Index verwendet werden würde? (2 Punkte)
 - Warum ist es nicht möglich zwei oder mehr Primärindexe oder Table cluster für eine Tabelle anzulegen? Erläutern Sie kurz ihre Antwort (2 Punkte)