

Datenbanken II

Übungsblatt 2 – WiSe 2019/20

4. Erstellen und zeichnen Sie für die folgende Tabelle einen 2-stufigen Sekundärindex auf dem Attribut Dept. D.h. der Index hat 2 Index-Stufen, wobei die erste Stufe dense und die zweite Stufe sparse sein soll. Pro Indexblock können 10 Einträge gespeichert werden. (Da es ein Sekundärindex ist, wird angenommen, dass ein Index-Eintrag auf ein Tupel verweist).

Name	Dept	CourseNo
Donetta	CS	457
Annabelle	Arch.	45
Roosevelt	CS	27
Allyson	Socio.	470
Debra	Psych.	457
Bobbie	Psych.	27
Bradly	CS	11
Marcell	CS	470
Amanda	Psych.	470
Danelle	Socio.	470
Michael	Arch.	125
Ann	Path.	350
Tom	Gen.	291
Camilla	Pol.	11
Abdul	CS	27
Carrie	CS	27
Ewa	Psych.	125
Conrad	CS	350
Lucille	Socio.	350
Roberto	Arch.	125

5. Ein Block kann 500 Index-Einträge oder 80 Datensätze der Relation R speichern. R enthält 10.000.000 Datensätze.
- Wieviele Blöcke werden für einen (flachen) *dense* Index auf R benötigt.
 - Wieviele Blöcke werden für einen (flachen) *sparse* Index auf R benötigt, der einen Eintrag pro Block der Daten Datei enthält.
6. Gegeben ist eine Relation $R[A, B, \dots]$ mit den folgenden Eigenschaften:
- $|R| = 1.000.000$ Tupel,
 - die Werte von Attribut A sind gleichverteilt im Intervall $[1, 100.000.000]$,
 - die Werte von Attribut B sind gleichverteilt im Intervall $[1, 1.000]$,
 - es existiert ein sparse Primärindex auf A (d.h. jeder Index-Eintrag verweist auf einen Datenblock von R),
 - es existiert ein dense Sekundärindex auf B (d.h. jeder Index-Eintrag verweist auf ein Tupel von R),
 - ein Block speichert 500 Index Einträge oder 50 Datensätze.

Es werden folgende Anfragen auf R gestellt:

$$Q1: \sigma_{A > 60.000.000}(R), \quad Q2: \sigma_{B > 600}(R)$$

Wie viele Blöcke müssen gelesen werden, wenn die Indizes zur Beantwortung von Q1 bzw. Q2 **nicht** verwendet werden.

7. Angabe wie in Aufgabe 6.
- Wie viele Blöcke müssen gelesen werden, wenn die Indizes zur Beantwortung von Q1 bzw. Q2 verwendet werden.
8. Gegeben ist eine Relation $R[A, B, C, E]$ mit den folgenden Eigenschaften:
- $|R| = 10.000.000$ Tupel
 - die Werte von Attribut A sind gleichverteilt im Intervall $[1, 100.000.000]$ und eindeutig,
 - die Werte von Attribut B sind gleichverteilt im Intervall $[1, 5.000]$,
 - die Werte von Attribut C sind gleichverteilt im Intervall $[1, 25.000]$,
 - die Werte von Attribut E sind gleichverteilt im Intervall $[1, 500.000]$,
 - es existiert ein flacher sparse Index auf Attribut A ,
 - auf den Attributen B, C und E existiert jeweils ein flacher dense Index,
 - ein Datenblock kann 25 Einträge speichern,
 - ein Indexblock kann 100 Einträge speichern.

Es werden die folgenden Anfragen gestellt:

$$Q1: \sigma_{C=20.000 \wedge A < 50.000.001}(R) \quad Q2: \sigma_{C=20.000 \vee A < 50.000.001}(R)$$

Beschreiben Sie die nötigen Schritte um Q1 bzw. Q2 möglichst effizient zu beantworten und geben Sie die notwendige Zahl an Blockzugriffe für die beschriebenen Schritte an.