

# Datenbanken II

## Übungsblatt 6 – WiSe 2016/17

18. Gegeben sei eine Relation  $R[A]$  mit einem Attribut  $A$ . Die folgenden 26 Werte sind (in gegebener Reihenfolge) in  $R$  gespeichert:

11, 25, 10, 0, 13, 9, 7, 22, 18, 5, 16, 3, 24, 20, 12, 8, 6, 2, 19, 14, 4, 21, 23, 17, 15, 1

Führen Sie externes Merge-Sort auf dieser Relation aus. Zeichnen Sie dafür den Schritt, in dem die Läufe erzeugt werden, sowie die einzelnen Merge-Schritte (vgl. Vorlesungsfolien zum Thema *Anfragebearbeitung* S. 16/52). Der Puffer hat eine Größe von  $M = 3$  Blöcken. Pro Block können 2 Tupel gespeichert werden.

Geben Sie außerdem die Anzahl der nötigen Blockzugriffe an.

19. Gegeben sei die (noch unsortierte) Relation  $R$  aus Aufgabe 18.

Führen Sie externes Merge-Sort auf  $R$  aus. Zeichnen Sie dafür den Schritt, in dem die Läufe erzeugt werden, sowie die einzelnen Merge-Schritte (vgl. wiederum Vorlesungsfolien zum Thema *Anfragebearbeitung* S. 16/52). Der Puffer hat diesmal eine Größe von  $M = 4$  Blöcken. Pro Block können wieder 2 Tupel gespeichert werden.

Geben Sie auch hier die Anzahl der nötigen Blockzugriffe an.

20. Gegeben sei eine Relation  $R[A, B]$ . Die Attributwerte von  $A$  und  $B$  sind positive ganze Zahlen. Es sind Duplikate erlaubt. Attribut  $A$  ist gleichverteilt im Wertebereich von  $[200.000; 240.000]$ , Attribut  $B$  ist gleichverteilt im Wertebereich von  $[400.000; 430.000]$ . Insgesamt sind in  $R$  10.000.000 Tupel gespeichert. Es gibt einen *clustered*  $B^+$  Baum Index auf dem Attribut  $A$  sowie einen *dense*  $B^+$  Baum Index auf dem Attribut  $B$ . Für beide  $B^+$  Bäume gilt  $m = 64$ . Die Größe eines Blockes beträgt 512 Bytes und die Tupelgröße beträgt 16 Bytes.

Schätzen Sie die Kosten für die folgende Anfrage ab:

$$Q1: \sigma_{A < 204.000 \vee A > 228.000} (R)$$

**Hinweis bzgl. clustered Index:** Wenn ein clustered Index auf einem Attribut existiert, sind die Blöcke der Daten-Datei bzgl. diesem Attribut sortiert. Pro Attributwert gibt es einen Eintrag im Index. Siehe auch Vorlesungsfolien zum Thema *Indexstrukturen* S. 10/89.

**Hinweis bzgl. des dense  $B^+$  Baum Indexes:** Bisher gab es keine Duplikate in  $B^+$  Bäumen. Um mit Duplikaten umzugehen gibt es mehrere Möglichkeiten. In dieser Aufgabe wird es folgendermaßen gelöst: Im Fall eines Duplikates wird ein Bucket für den betroffenen Schlüssel angelegt, in dem alle Verweise auf die jeweiligen Datenblöcke gespeichert werden (vgl.

bspw. Vorlesungsfolien zum Thema *Indexstrukturen* S.14/89, Key 700 in der Abbildung). Ein solcher Bucket ist 512 Bytes groß und ein Eintrag benötigt 8 Bytes. Werden mehrere Buckets für einen Wert benötigt, dann werden die einzelnen Buckets in einer verketteten Liste gespeichert. Im Falle einer Verkettung wird der letzte Eintrag im Bucket für den Verweis auf den nächsten Bucket verwendet.

21. Angabe, bis auf die abzuschätzende Anfrage, wie in Aufgabe 20.  
Schätzen Sie die Kosten für die folgende Anfrage ab:

$$\text{Q2: } \sigma_{A < 208.000 \wedge B > 415.000} (R)$$

22. Angabe, bis auf die abzuschätzende Anfrage, wie in Aufgabe 20.  
Schätzen Sie die Kosten für die folgende Anfrage ab:

$$\text{Q3: } \sigma_{A = 228.000 \wedge B < 409.000} (R)$$

23. Angabe, bis auf die abzuschätzende Anfrage, wie in Aufgabe 20.  
Schätzen Sie die Kosten für die folgende Anfrage ab:

$$\text{Q4: } \sigma_{A < 228.000 \wedge B = 409.000} (R)$$