

# Datenbanken II

## Übungsblatt 8 – WiSe 2017/18

20. [*Externes Merge-Sort*]

Gegeben sei eine (unsortierte) Relation  $R$  mit 26 Tupeln.

Sie wollen externes Merge-Sort auf  $R$  durchführen. Pro Block können 2 Tupel gespeichert werden. Anfangs steht Ihnen ein Puffer der Größe  $M = 3$  Blöcke zur Verfügung.

Sie wollen, dass bei der Durchführung des Merge-Sort-Algorithmus nur ein einziger **Merge-Schritt** nötig ist. Dies funktioniert mit der gegebenen Puffergröße leider nicht.

Um das Problem zu lösen könnte man den Puffer unendlich groß machen. Sie wollen aber auch nicht unnötig Speicher verschwenden.

Geben Sie die **minimale Puffergröße** (in Blöcken) an, sodass der Merge-Sort-Algorithmus nur einen einzigen **Merge-Schritt** benötigt und **begründen** Sie Ihre Antwort.

**Hinweis:** Sie können diese Aufgabe auch durch Probieren lösen, aber Ihre Begründung sollte trotzdem schlüssig sein.

21. [*Selektionsauswertung*]

Gegeben sei eine Relation  $R[A, B]$  mit folgenden Eigenschaften:

- Die Attributwerte von  $A$  und  $B$  sind positive ganze Zahlen.
- Es sind Duplikate erlaubt.
- Attribut  $A$  ist gleichverteilt im Wertebereich  $[200.001; 231.250]$ .
- Attribut  $B$  ist gleichverteilt im Wertebereich  $[300.001; 400.000]$ .
- Insgesamt sind in  $|R| = 10.000.000$  Tupel gespeichert.
- Es existiert ein *clustered*  $B^+$  Baum Index auf dem Attribut  $A$ .
- Es existiert ein *dense*  $B^+$  Baum Index auf dem Attribut  $B$ .
- Für beide  $B^+$  Bäume gilt:
  - Knotengrad  $m = 2^8$ .
  - Duplikate werden per Tuple Identifier (TID) eindeutig gemacht.
  - Es ist nichts über den Füllgrad der Knoten bekannt, d.h. Sie müssen vom schlechtesten Fall ausgehen (alle Knoten sind nur halb voll).
- Blockgröße:  $2^{10}$  Bytes, Tupelgröße:  $2^4$  Bytes.

Geben Sie die **Strategie** und die **Kosten** zur Auswertung von Q1 an:

$$Q1: \sigma_{A < 206.251 \vee A > 228.000} (R)$$

22. [Bitmap Index Scan]

Gegeben sei die Relation  $R[Name, S]$  in Abbildung 1. In  $R$  sind 15 Tupel auf 5 Blöcken gespeichert (Block Adr. 0 - 4). Weiters existiert ein flacher dense Index auf dem Attribut  $S$  (vgl. ebenfalls Abbildung 1).

*Annahme:* Der gesamte Index passt auf einen Block.

Die folgende Anfrage soll unter Verwendung des Indexes beantwortet werden:

$$Q2: \sigma_{S \geq 10 \wedge S < 600}(R)$$

- a) Es wird **kein** Bitmap Index Scan verwendet um Q2 zu beantworten. Geben Sie die Blockzugriffe an, bspw. 1 Indexblock, Datenblöcke: 2, 3, 0, 2, ...).
- b) Es wird ein Bitmap Index Scan verwendet um Q2 zu beantworten (siehe S. 27/52 in den Vorlesungsfolien zum Thema *Anfragebearbeitung*<sup>1</sup>).
  - i. Geben Sie die Länge des Bit-Arrays an.
  - ii. **Visualisieren** Sie, wie sich das Bit-Array verändert.
  - iii. Geben Sie wieder die Blockzugriffe an und **beschreiben** Sie anhand der Aufgabe, **warum** ein Bitmap Index Scan die Anfrageauswertung effizienter machen kann (im Vergleich zur Auswertung mit Index, aber ohne Bitmap Index Scan).

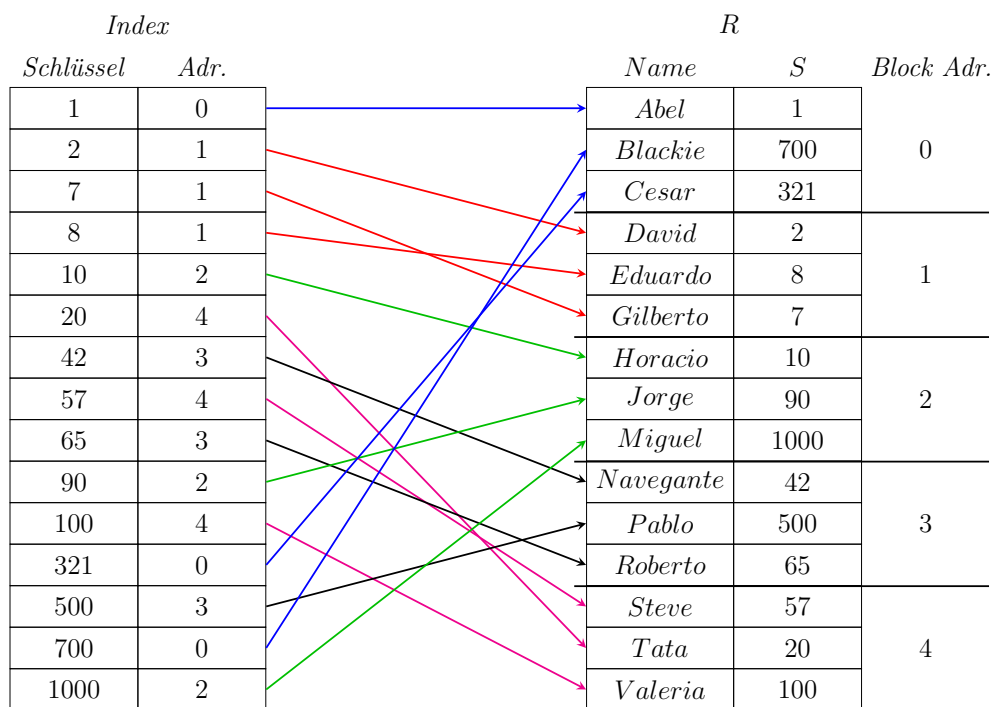


Abbildung 1: Relation  $R[Name, S]$  und dense Index für Aufgabe 22.

<sup>1</sup>[https://dbresearch.uni-salzburg.at/teaching/2017ws/db2/db2\\_03-handout-1x1-prev.pdf](https://dbresearch.uni-salzburg.at/teaching/2017ws/db2/db2_03-handout-1x1-prev.pdf)