

Aufgabe 1

1 Punkt

Betrachten Sie die folgende Festplatte:

- Sektorgröße $B = 2^9\text{B}$,
- Interblock Gap Size $G = 0\text{B}$,
- Sektoren/Spur $S = 2^8$,
- Spuren pro Scheibenseite $T = 300$,
- Anzahl der **beidseitig** beschriebenen Scheiben $D = 2^3$,
- mittlerer Spurwechsel $sp = 2\text{ms}$,
- Drehzahl $dz = 60 * 2^7\text{rpm}$.

Wie hoch ist die maximale Datenrate (MB/s) für sequentielles Lesen auf dieser Disk, wenn von allen Seiten/Scheiben gleichzeitig gelesen wird?

Lösung: _____

Aufgabe 21 Punkt

In einer Slotted Page der Größe 8192 Bytes sind 200 Datensätze gespeichert. Was ist die maximale Größe eines Datensatzes, wenn alle Datensätze gleich groß sind? Der Header ist dense. Es wird Byte-Adressierung verwendet.

Lösung: _____

Aufgabe 3

1 Punkt

Gegeben ist eine Relation $R[A, B, C, D]$. R ist in einer sequentiellen Datei nach Attribut B sortiert gespeichert und enthält 180.000 Tupel. Die Werte der Attribute A und B sind eindeutig und liegen zwischen 1 und 1.000.000. Jeder Datenblock kann bis zu 12 Tupel enthalten. Es gibt einen einstufigen dense Index $Index_1$ (**kein** B⁺-Baum Index) auf Attribut A . Weiters gibt es einen einstufigen sparse Index $Index_2$ (**kein** B⁺-Baum Index) auf $Index_1$. Auf dem Attribut B gibt es einen sparse B⁺-Baum Index. Jeder Indexblock kann bis zu 60 Einträge (Key + Pointer) beinhalten (für B⁺-Baum: Anzahl Pointer pro Block ist 60).

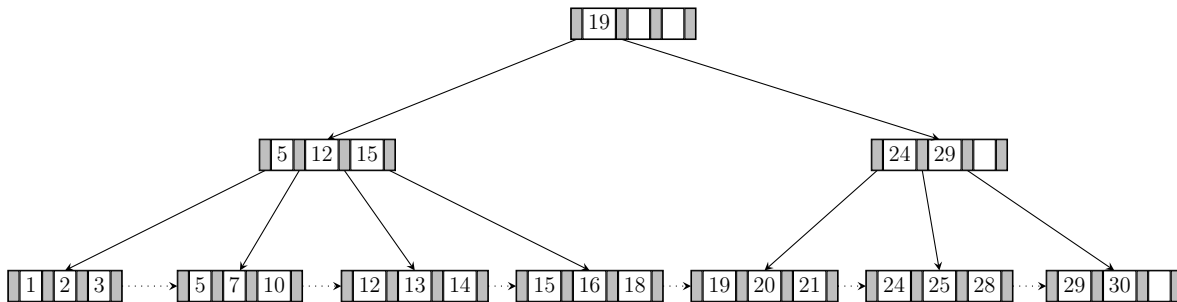
Geben Sie die Strategie sowie die Anzahl der nötigen Blockzugriffe für die effizienteste Zugriffsmethode an, um die Anfrage $\sigma_{A=50000}(R)$ zu beantworten.

Lösung: _____

Aufgabe 4

1 Punkt

Gegeben ist der unten angeführte B+-Baum mit $m = 4$. Fügen Sie den Wert 8 gemäß des in der Vorlesung besprochenen Algorithmus in diesen B+-Baum ein.



Lösung: _____

Aufgabe 5

1 Punkt

Die folgende Tabelle wird in eine statische Hash-Datei überführt. Als Suchschlüssel wird das Attribut *persNum* verwendet. Der Hashwert ist die letzte Ziffer des Attributwerts. Nehmen Sie an, dass kein Bucket Overflow auftritt. Illustrieren Sie den Inhalt der daraus resultierenden Hash-Datei.

<i>Name</i>	<i>persNum</i>
Uriah	2941
Carly	1564
Iliana	9050
Fredericka	8901
Ira	7297
Moana	5181
Daniel	9265
Octavius	7990
Denton	3622
Mia	8287
Curran	6020

Name:

Matrikelnummer:

7/13

Aufgabe 6

1 Punkt

Führen Sie externes Merge-Sort auf der Relation $R[A]$ aus. Die Werte des Attributs A der Relation sind

$A = (36, 9, 84, 78, 66, 27, 18, 45, 0, 75, 12, 48, 24, 81, 3, 69, 30, 39, 60, 72, 42, 21, 87, 33, 51, 90, 15, 57, 54, 63, 6)$

in der angegebenen Reihenfolge. Im Puffer stehen 5 Blöcke zur Verfügung. Jeder Block fasst 2 Tupel. Geben Sie die Anzahl der notwendigen Blockzugriffe an.

Aufgabe 7**1 Punkt**

Gegeben ist die Relation $R[A]$. Auf $R.A$ existiert ein sparse Index. Was ist die effizienteste Strategie um Bereichsanfragen von folgendem Typ zu beantworten?

$$\sigma_{a < A < b}(R)$$

Geben Sie alle notwendigen Schritte an.

Aufgabe 81 Punkt

Gegeben ist eine Relation $R[A, B]$. R beinhaltet 100000 Tupel, wobei es 50 verschiedene Werte von Attribut B (gleichverteilt) gibt. Die Relation R ist auf 200 Blöcken gespeichert. Es gibt einen dense B⁺-Baum Index auf Attribut B . Jeder Knoten im B⁺-Baum kann maximal 100 Schlüssel beinhalten. Die Duplikate werden durch Verknüpfung mit dem TID (Tuple Identifier) eindeutig gemacht. Es soll folgende Anfrage ausgeführt werden:

$$\sigma_{B=b}(R)$$

Erklären Sie wie man einen Bitmap Index Scan in diesem Szenario benutzen kann und beantworten Sie die Anfrage möglichst effizient.

Aufgabe 9

1 Punkt

Berechnen Sie die Kosten (Blockzugriffe) von einem Natural Join zwischen zwei Relationen $R[A, B, C]$ und $S[B, C, D]$, wobei $|R| = 10000$ Tupel und $|S| = 20000$ Tupel. Die Relationen sind auf 2000 bzw. 5000 hintereinander liegenden Blöcken gespeichert. Es existiert ein dense B^+ -Baum Index auf $R.B$ und auf $S.(B, C)$ (ein Schlüssel ist die Kombination von beiden Attributen, B und C , in dieser Reihenfolge). Die Attributwerte von $R.(B, C)$ und auf $S.(B, C)$ sind eindeutig. Jeder Knoten im B^+ -Baum 100 kann Schlüssel speichern. Wenn möglich soll der Join mit Hilfe des Index durchgeführt werden.

Lösung: _____

Aufgabe 101 Punkt

Welcher Join Algorithmus (Sort-Merge Join, Block Nested Loop Join) generiert die minimalen Kosten für das folgende Szenario? Geben Sie in der Lösung den Namen des Algorithmus **und** die dazugehörigen Kosten an.

Berechnen Sie einen Natural Join zwischen zwei Relationen $R[A, B]$ und $S[B, C]$, wobei $|R| = 100000$ Tupel und $|S| = 20000$ Tupel. Die Relationen sind auf $b_R = 10000$ bzw. $b_S = 10000$ hintereinander liegenden Blöcken gespeichert. Der Buffer hat Platz für 51 Blöcke. Es existiert ein sparse B⁺-Baum Index auf $S.B$ und ein dense B⁺-Baum Index auf $R.B$. Jeder Knoten im B⁺-Baum kann 100 Schlüssel speichern.

Lösung: _____

Aufgabe 11**1 Punkt**

Gegeben sind zwei Relationen $R[A, B, \underline{C}, \underline{D}]$ und $S[\underline{C}, \underline{D}, E, F]$ (d.h., das Paar (C,D) ist ein Schlüssel in R und S).

Formulieren Sie die folgende Anfrage so um, dass Selektionen und Projektionen so weit wie möglich nach unten geschoben werden (sodass der kleinstmögliche Join durchgeführt wird). Zeichnen Sie zuerst die Normalform.

$$\pi_A (\sigma_{D=d}(R) \bowtie \sigma_{C=c \wedge E=e}(S))$$

Aufgabe 12

1 Punkt

Gegeben sind folgende Relationen:

- $|R[A, B, C]| = 1000, V(R, A) = 100, V(R, B) = 200, V(R, C) = 300$
- $|S[A, D, E]| = 4000, V(S, A) = 50, V(S, D) = 200, V(S, E) = 300$
- $|T[D, E, F]| = 2000, V(T, D) = 200, V(T, E) = 400, V(T, F) = 600$

Die Werte in den Tupeln sind gleichverteilt und unabhängig.

Schätzen Sie die Kardinalität des Ergebnisses der folgenden Anfrage ab ($\sigma_{A=100}(R) \neq \emptyset$).

$$(\sigma_{A=100}(R)) \bowtie S \bowtie T$$

Comments:

- **[TODO]** Expand the solution to the full formula (with mins not maxs).

Lösung: _____