

# Datenbanken II

## Übungsblatt 1 – WiSe 2014/15

1. Betrachte folgende Festplatte: Sektorgröße  $B = 512\text{B}$ , Interblock Gap Size  $G = 64\text{B}$ , Sektoren/Spur  $S = 100$ , Spuren pro Scheibenseite  $T = 500$ , Anzahl der beidseitig beschriebenen Scheiben  $D = 20$ , mittlerer Spurwechsel  $sp = 5\text{ms}$ , Drehzahl  $dz = 5400\text{rpm}$ .

Bestimme die folgenden Werte:

- Bruttokapazität pro Spur,  $bks$
  - nutzbare Kapazität pro Spur,  $nks$
  - Anzahl der Zylinder
  - nutzbare Kapazität pro Zylinder,  $nkz$
  - nutzbare Kapazität der Festplatte,  $nkf$
  - Zugriffszeit pro Sektor,  $tz$
  - Datenrate,  $dr$
  - Zeit für das Lesen von 20 nicht-sequentiellen Sektoren,  $tr20$
  - Zeit für das Lesen von 20 sequentiellen Sektoren,  $ts20$
2. Betrachte eine Festplatte mit 200 Zylinder. Auf die folgenden Spuren soll zugegriffen werden: 102, 167, 61, 143, 16, 159, 81, 97. Der Kopf ist über der Spur 75 positioniert. Bestimme die Reihenfolge, in der die Spuren gelesen werden, und die gesamte Armbewegungskosten (d.h. die Anzahl der Sektoren über die der Kopf bewegt wird) für die folgenden Disk Scheduling Algorithmen.
- FCFS: First Come First Serve. Die Anfragen werden in der Reihenfolge bearbeitet, in der sie bei der Festplatte eintreffen.
  - SSTF: Shortest Seek Time First. Die Anfragen werden nach der Suchzeit, die der Kopf benötigt, sortiert und abgearbeitet. Die Suchzeit ist jene Zeit, die der Kopf benötigt, um von einer Spur zur nächsten zu gelangen. Bspw. wenn der Kopf von Spur 75 auf Spur 81 wechselt, ist die Seek Time =  $|75 - 81| = 6$ . Anmerkung: Bei gleicher Seek Time entscheidet der Zufall.
  - LOOK. Hier wird nach dem Elevator-Prinzip gearbeitet: Es wird zuerst in eine Richtung gescannt. Diese Richtung wird frei gewählt aber muss danach beibehalten werden bis der Lesekopf *bei der letzten zu lesenden Spur dieser Richtung* angelangt ist. Bspw. wenn der Kopf von Spur 75 Richtung 0 scannt, werden alle Anfragen auf dem Weg zwischen 75 und 0 abgearbeitet. Genauso könnte vom Spur 75 Richtung 199 gescannt werden (wie in der Lösung unten). Wenn die letzte Spur der Richtung

erreicht ist, wird die Richtung geändert und nach demselben Prinzip gearbeitet.

3. Gegeben zwei Relationen  $R(A)$  und  $S(A)$ . Die Werte in  $R$  sind nicht sortiert,  $S$  ist nach dem Attribut  $A$  sortiert.  $R$  und  $S$  speichern dieselben numerischen Werte, die zwischen 5.000.000 und 10.000.000 gleichverteilt sind; ein bestimmter Wert kann auch mehrfach vorkommen.

Blockgröße  $B = 8192\text{B}$ . Tupelgröße  $t = 120\text{B}$ .  $n = |R| = |S| = 1.000.000$  Tupel. Die Zeit für 1 Lesezugriff auf einen Block ist  $0.02\text{s}$ .

Ermittle die Ausführzeit für folgende Anfragen, wobei entweder  $X = R$  oder  $X = S$ .

- a)  $\sigma_{A=7.000.000}(X)$
  - b)  $\sigma_{A<7.000.007}(X)$
  - c)  $\sigma_{A\neq 7.000.000}(X)$
4. Gegeben zwei Relationen,  $R$  und  $S$ , so dass  $|R| = |S| = 1000$ . Blockgröße  $B = 8192\text{B}$ , Tupelgröße  $t = 96\text{B}$ , Puffergröße  $p = 90\text{kB}$ . Berechne wie viele Blockzugriffe ein Nested-Loop Join mit den folgenden Strategien braucht.
    - a) LRU
    - b) MRU