



Übung zur Vorlesung
Einsatz und Realisierung von Datenbanksystemen im SoSe16

Moritz Kaufmann (moritz.kaufmann@tum.de)
<http://db.in.tum.de/teaching/ss16/impldb/>

Blatt Nr. 07

Hausaufgabe 1

Geben die Relation Klausur:

MatrNr	Vorbereitungszeit	Note
1	150	1.7
2	70	2.7
3	450	2.0
4	180	1.7
5	2500	1.3

- Formulieren Sie die Anfrage, die die MatrNr in der Skyline für die Attribute Vorbereitungszeit und Note erzeugt (kleiner ist jeweils besser) in SQL mit Hilfe des Skyline Operators.
- Formulieren Sie die Anfrage in SQL ohne Skyline Operator.
- Bestimmen Sie das Ergebnis der Anfrage.

Hausaufgabe 2

Folgende Datenpunkte im euklidischen Raum seien gegeben:

Punkt	X	Y
A	2	12
B	3	11
C	3	8
D	5	4
E	7	5
F	7	3
G	10	8
H	13	8

Clustern Sie die Punkte mithilfe des *k-means*-Verfahren in 3 Cluster. Nutzen Sie als initiale Clusterzentren die Werte *A*, *B* und *C*. Wenn ein Punkt zu mehreren Clustern die gleiche Distanz hat, wird er dem Cluster der näher am Nullpunkt liegt zugeordnet. Geben Sie für jede Iteration jeweils die Zuordnung und die Mittelpunkte der Cluster an.

Hausaufgabe 3

In Hauptspeicherdatenbanken ist die Geschwindigkeit oft durch Limitierungen des Speichersystems begrenzt. Analysieren sie dazu folgende Fragestellungen:

1. Was versteht man unter NUMA und welche Schichten gibt es in der Speicherhierarchie? Geben Sie zu jeder Schicht auch die Zugriffszeiten und Bandbreite an.
2. Was bedeuten die Begriffe *Cacheline* und *Seite*. Auf welcher Schicht sind diese jeweils relevant?

Hausaufgabe 4

Gegeben eine Tabelle *Produkte* mit folgendem Schema und 10000 Einträgen:

Id (8 Byte) | Name (32 Byte) | Preis (8 Byte) | Anzahl (8 Byte)

Wieviele Daten werden für folgende Queries in die CPU-Caches geladen? Unterscheiden sie jeweils zwischen Row und Column Store.

1. *select * from Produkte*
2. *select Anzahl from Produkte*