



Übung zur Vorlesung
Einsatz und Realisierung von Datenbanksystemen im SoSe15

Moritz Kaufmann (moritz.kaufmann@tum.de)
<http://db.in.tum.de/teaching/ss15/impldb/>

Blatt Nr. 10

Hausaufgabe 1

Sie eine SQL-Anfrage, die basierend auf dem Schema aus Abbildung 1 einen dreidimensionalen Quader berechnet, der es unserem Handelsunternehmen erlaubt, entlang der folgenden Dimensionen drill-down/roll-up Anfragen zu stellen:

- Produkttyp,
- Bezirk,
- Alter.

Das Handelsunternehmen ist dabei nur an Daten aus Deutschland interessiert, die in die Hochsommersaison fallen. Verwenden Sie den **cube**-Operator.

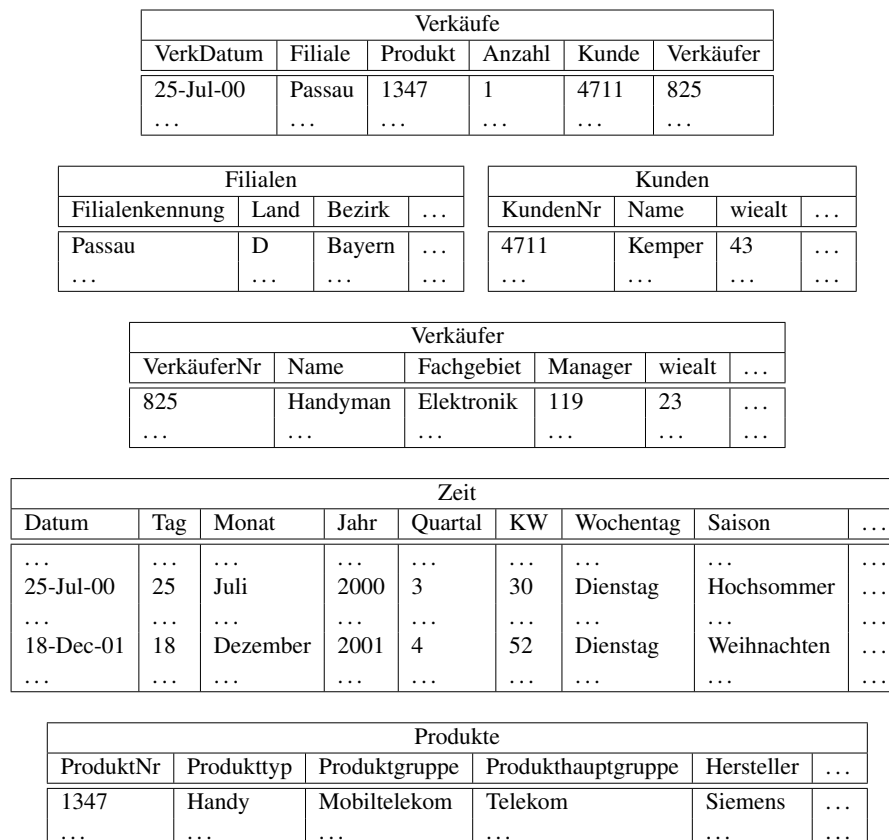


Abbildung 1: Schema des Handelsunternehmens.

Hausaufgabe 2 - Top-K Berechnung

Die in Abbildung 2 dargestellten Relationen Autos und Unterhalt dienen der Bewertung von Autos. Eine junge Studierende sucht ein Auto mit guter Balance zwischen Sportlichkeit und Kosten. Sie überlegt sich wie die drei Werte Preis, PS und monatlicher Unterhalt in einen Score umberechnet werden können und nutzt schließlich folgende Formel:

$$\text{Preis} - (100 * PS) + 24 * \text{Unterhalt}$$

Zeigen Sie die phasenweise Berechnung der Top-3 Ergebnisse jeweils mit dem Threshold- und dem NRA-Algorithmus. Prüfen sie vor der Berechnung ob Teile der Scoringformel schon innerhalb jeder Relation vorberechnet werden können.

Auto	Preis	PS	Auto	Unterhalt p. Monat
Seat Leon	25000€	200	Seat Leon	215€
Audi A1	17000€	96	Audi A1	220€
Citroen DS 4	20679€	100	Citroen DS 4	225€
Mini One	16500€	75	Mini One	262€
Mercedes C-Klasse	35000€	160	Mercedes C-Klasse	290€
Porsche Cayenne	80100€	420	Porsche Cayenne	430€

Abbildung 2: Autokauf und -Unterhaltskosten.

Hausaufgabe 3 - Skyline

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 4 - Frequent Itemsets

Zeigen Sie die weiteren Phasen des \dot{A} priori-Algorithmus für unser Beispiel in Abbildung 3 (hier ist lediglich bis inkl. 2. Phase dargestellt). Damit eine Menge von Produkten ein frequent itemset ist, muss sie in mindestens $3/5$ aller Verkäufe enthalten sein, d.h. $minsupp = 3/5$.

VerkaufsTransaktionen		Zwischenergebnisse	
TransID	Produkt	FI-Kandidat	Anzahl
111	Drucker	{Drucker}	4
111	Papier	{Papier}	3
111	PC	{PC}	4
111	Toner	{Scanner}	2
222	PC	{Toner}	3
222	Scanner	{Drucker, Papier}	3
333	Drucker	{Drucker, PC}	3
333	Papier	{Drucker, Scanner}	3
333	Toner	{Drucker, Toner}	3
444	Drucker	{Papier, PC}	2
444	PC	{Papier, Scanner}	3
555	Drucker	{Papier, Toner}	3
555	Papier	{PC, Scanner}	2
555	PC	{PC, Toner}	2
555	Scanner	{Scanner, Toner}	2
555	Toner		

Abbildung 3: Ausgangssituation für den \dot{A} priori-Algorithmus

Hausaufgabe 5 - Row vs Column - Store

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 den CPU-Cache geladen? Unterscheiden sie jeweils zwischen Row und Column Store.

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