



Übung zur Vorlesung *Einsatz und Realisierung von Datenbanksystemen* im SoSe17

Maximilian E. Schüle (schuele@in.tum.de)
<http://db.in.tum.de/teaching/ss17/impldb/>

Blatt Nr. 07

Hausaufgabe 1 Cubes mit Studenten

Erstellen Sie anhand des Universitätsschemas einen Cube mit der Faktentabelle *pruefen* und den Dimensionstabellen *Vorlesungen*, *Professoren*, *Studenten*. Zählen Sie das Auftreten von Vorlesungen, Professoren und Studenten.

```
with mehr_pruefen(MatrnNr,VorlNr,PersNr>Note) as (  
    select * from pruefen  
    union  
    values (28106,5041,2137,3.0)  
)
```

Hausaufgabe 2

Schreiben 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.

Hausaufgabe 3 Fensterln mit SQL

Analysieren wir die Gehälter von Professoren mittels Windowfunctions und führen Sie die Abfragen unter hyper-db.de aus. Dazu orientieren wir uns an der Relation *Professoren* des erweiterten Universitätsschemas:

```
with Professoren (persnr, name, rang, raum, gehalt, steuerklasse) as (  
    values (2125,'Sokrates','C4',226,85000,1) UNION  
    values (2126,'Russel','C4',232,80000,3) UNION  
    values (2127,'Kopernikus','C3',310,65000,5) UNION  
    values (2128,'Aristoteles','C4',250,85000,1) UNION  
    values (2133,'Popper','C3',52,68000,1) UNION  
    values (2134,'Augustinus','C3',309,55000,5) UNION  
    values (2136,'Curie','C4',36,95000,3) UNION  
    values (2137,'Kant','C4',7,98000,1)  
)
```

Verkäufe					
VerkDatum	Filiale	Produkt	Anzahl	Kunde	Verkäufer
25-Jul-00	Passau	1347	1	4711	825
...

Filialen				Kunden			
Filialenkennung	Land	Bezirk	...	KundenNr	Name	wiealt	...
Passau	D	Bayern	...	4711	Kemper	43	...
...

Verkäufer					
VerkäuferNr	Name	Fachgebiet	Manager	wiealt	...
825	Handyman	Elektronik	119	23	...
...

Zeit								
Datum	Tag	Monat	Jahr	Quartal	KW	Wochentag	Saison	...
...
25-Jul-00	25	Juli	2000	3	30	Dienstag	Hochsommer	...
...
18-Dec-01	18	Dezember	2001	4	52	Dienstag	Weihnachten	...
...

Produkte					
ProduktNr	Produkttyp	Produktgruppe	Produkthauptgruppe	Hersteller	...
1347	Handy	Mobiltelekom	Telekom	Siemens	...
...

Abbildung 1: Schema des Handelsunternehmens.

1. Ermitteln Sie zu jedem Professor das Durchschnittsgehalt aller Professoren.
2. Ermitteln Sie zu jedem Professor das Durchschnittsgehalt aller Professoren partitioniert nach Rang.
3. Ermitteln Sie nun die wachsende Summe (das Quantil) des Gehaltes aller Professoren partitioniert nach Rang und absteigend sortiert nach ihrem Gehalt. Gleich verdienende Professoren sind im selben Quartil.
4. Ermitteln Sie nun die wachsende Summe des Gehaltes aller Professoren partitioniert nach Rang und absteigend (total) sortiert nach ihrem Gehalt (reihenweise, nicht als Range-Query).
5. Ermitteln Sie nun das gleitende Durchschnittsgehalt aus genau zwei mehr bzw. weniger verdienenden Professoren sortiert nach Gehalt und partitioniert nach Rang.
6. Ermitteln Sie nun das gleitende Durchschnittsgehalt aus den 500 Einheiten mehr bzw. weniger verdienenden Professoren sortiert nach Gehalt und partitioniert nach Rang. (weder in PostgreSQL noch in HyPer implementiert, fragen Sie Ihren Tutor)
7. Geben sie zu jedem Professor das Gehalt des eins besser wie eins schlechter verdienenden.
8. Ermitteln Sie die drei bestverdienendsten Professoren einmal mit und einmal ohne Windowfunctions.

Hausaufgabe 4 Lösen sie folgende Anfrage mit SQL basierend auf dem bekannten Universitätschema.

1. Bestimmen Sie die Durchschnittsnote für jeden Studenten.
2. Basierend auf dieser Durchschnittsnote, bestimmen Sie für alle Studenten ihren Rangplatz innerhalb ihrer Kohorte (Studenten desselben Semesters).
3. Berechnen Sie zusätzlich für jeden Studenten auch noch die ABWEICHUNG seiner Durchschnittsnote von der Durchschnittsnote der Kohorte (also vom Durchschnitt der Durchschnittsnote der Studenten der Kohorte) ausgegeben werden.

Lösen Sie Teilaufgaben 2 und 3 jeweils einmal mit und einmal ohne Nutzung von Windowfunktionen. Ihre Anfragen können Sie auf hyper-db.de testen. Nutzen sie folgende erweiterte *pruefen* Relation:

```
with mehr_pruefen(MatrnR,VorlNr,PersNr>Note) as (
  select * from pruefen
  union
  values (29120,0,0,3.0), (29555,0,0,2.0), (29555,0,0,1.3), (29555,0,0,1.0)
)
```

Hausaufgabe 5 Germany - bad, very bad!

Ermitteln Sie den Handelsüberschuss zwischen Deutschland und den USA auf Basis des Schneeflocken-Schemas des TPC-H-Benchmarks. Orientieren Sie sich an der TPC-H-Abfrage 7 und nutzen Sie hyper-db.de.

Hausaufgabe 6

Berechnen Sie, wie groß ein Data Warehouse für ein Handelsunternehmen wie Zalando oder Amazon.com wäre, wenn die Bestelldaten der letzten 3 Jahre enthalten sind. Versuchen sie ihre Schätzung basierend auf öffentlichen Daten wie Umsatz o.Ä. durchzuführen.