TU München, Fakultät für Informatik
Lehrstuhl III: Datenbanksysteme
Prof. Alfons Kemper, Ph.D.

Übung zur Vorlesung *Einsatz und Realisierung von Datenbanken* im SoSe20

Maximilian {Bandle, Schüle}, Josef Schmeißer (i3erdb@in.tum.de)
http://db.in.tum.de/teaching/ss20/impldb/

Blatt Nr. 06

Hausaufgabe 1

Gehen Sie von folgender kombinierter Fragmentierung der in Abbildung 1 dargestellten Relation *Professoren* aus:

Professoren						
PersNr	Name	Rang	Raum	Fakultät	Gehalt	Steuerklasse
2125	Sokrates	C4	226	Philosophie	85000	1
2126	Russel	C4	232	Philosophie	80000	3
2127	Kopernikus	C3	310	Physik	65000	5
2133	Popper	C3	52	Philosophie	68000	1
2134	Augustinus	C3	309	Theologie	55000	5
2136	Curie	C4	36	Physik	95000	3
2137	Kant	C4	7	Philosophie	98000	1

Abbildung 1: Beispielausprägung der um drei Attribute erweiterten Relation *Professoren*

1. Zuerst erfolgt eine vertikale Fragmentierung in

$$\begin{aligned} \text{ProfVerw} &:= \Pi_{\text{PersNr, Name, Gehalt, Steuerklasse}}(\text{Professoren}) \\ \text{Profs} &:= \Pi_{\text{PersNr, Name, Rang, Raum, Fakultät}}(\text{Professoren}) \end{aligned}$$

2. Das Fragment Profs wird weiter horizontal fragmentiert in

$$\begin{aligned} \text{TheolProfs} &:= \sigma_{\text{Fakultät = 'Theologie'}}(\text{Profs}) \\ \text{PhysikProfs} &:= \sigma_{\text{Fakultät = 'Physik'}}(\text{Profs}) \\ \text{PhiloProfs} &:= \sigma_{\text{Fakultät = 'Philosophie'}}(\text{Profs}) \end{aligned}$$

Übersetzen Sie aufbauend auf dieser Fragmentierung die folgende SQL-Anfrage in die kanonische Form.

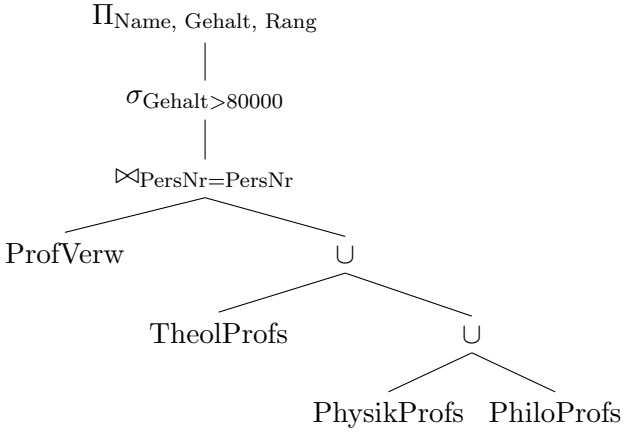
```
select Name, Gehalt Rang
from Professoren
where Gehalt > 80000;
```

Optimieren Sie diesen kanonischen Auswertungsplan durch Anwendung algebraischer Transformationsregeln (Äquivalenzen).

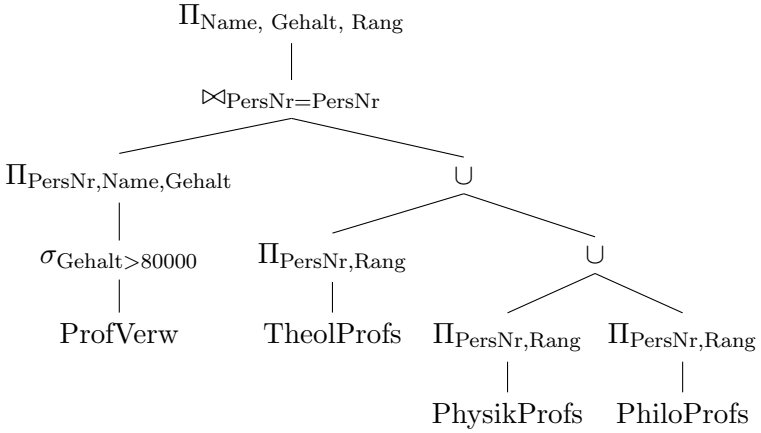
Vgl. Übungsbuch.

$$\Pi_{\text{Name, Gehalt, Rang}}(\Pi_{\text{PersNr,Name, Gehalt}}(\sigma_{\text{Gehalt} > 80000}(\text{ProfVerw}))\bowtie_{\text{PersNr=PersNr}} (\Pi_{\text{PersNr, Rang}}(\text{TheolProfs}) \cup \Pi_{\text{PersNr, Rang}}(\text{PhysikProfs}) \cup \Pi_{\text{PersNr, Rang}}(\text{PhiloProfs})))$$

Der Baum zum kanonischen Auswertungsplan sieht wie folgt aus:



In einem ersten Schritt verschiebt man die Selektion näher an die Datenquellen, die sich nur auf *ProfVerw* bezieht. Anschließend versuchen wir, die Zwischenergebnisse so klein wie möglich zu halten, indem wir zusätzliche Projektionen einfügen. Das Attribut *Name* ist redundant in beiden vertikalen Fragmenten enthalten und wird nicht benötigt.



Hausaufgabe 2

Für die Rekonstruierbarkeit der Originalrelation R aus vertikalen Fragmenten $R_1, \ldots, R_n$ reicht es eigentlich, wenn Fragmente paarweise einen Schlüsselkandidaten enthalten. Illustrieren Sie, warum es also nicht notwendig ist, dass der Durchschnitt aller Fragmentschemata einen Schlüsselkandidaten enthält. Es muss also nicht unbedingt gelten

$$R_1 \cap \cdots \cap R_n \supseteq \kappa,$$

wobei κ ein Schlüsselkandidat aus R ist.

Geben Sie ein anschauliches Beispiel hierfür – am besten bezogen auf unsere Beispiel-Relation *Professoren*.

Lsg: Vgl. Übungsbuch, Primärschlüssel unterstrichen: RaumF: $\{[\underline{Raum}, Fakultaet]\}$
Professoren: $\{[\underline{PersNr}, Name, Raum]\}$, ProfessorenR: $\{[\underline{PersNr}, Rang]\}$
ProfessorenF: $\{[\underline{PersNr}, Name, Rang, Raum, Fakultaet]\}$
ProfessorenF = RaumF $\bowtie_{\text{Raum=Raum}}$ (Professoren $\bowtie_{\text{PersNr=PersNr}}$ (ProfessorenR))

Hausaufgabe 3

Gegeben sei folgende Relation `Klausur` mit Schlüssel `MatrNr`:

MatrNr	Name	Note	Standort
10101	Philipp	1,0	München
10102	Magdalena	1,0	Garching
10103	Erik	1,0	Garching
10104	Josef	1,0	Garching
10105	Alex	1,0	Garching
10106	Maxmilian	1,0	München

Für eine verteilte Datenbank soll die Tabelle geeignet fragmentiert werden. Ziel ist, Namen mit Standort der Studenten lokal und die Noten getrennt abzuspeichern.

1) Fragmentieren Sie die Relation geeignet *vertikal*.
   a) Geben Sie das Schema für die zwei resultierenden Relationen $KlausurV_1$ und $KlausurV_2$ an. Unterstreichen Sie jeweils den Primärschlüssel.

      $KlausurV_1 : \{[\underline{MatrNr}, Note]\}, KlausurV_2\{[\underline{MatrNr}, Name, Standort]\}$

   b) Geben Sie in SQL-92 die zwei resultierenden Relationen `KlausurV1` und `KlausurV2` als Hilfstabellen (mittels `with`) an.

```
with KlausurV1 as (SELECT MatrNr,Note FROM Personen),
     KlausurV2 as (SELECT MatrNr,Name,Standort FROM Personen)
```

2) Die geeignetere der beiden resultierenden Relationen soll *horizontal* fragmentiert werden.
   a) Geben Sie das Prädikat der Selektion an, mit dem fragmentiert wird.

      Standort='Garching' oder Standort='München'

   b) Geben Sie in SQL-92 die zwei resultierenden Relationen `KlausurH1` und `KlausurH2` als Hilfstabellen (mittels `with`) an.

```
with KlausurH1 as(SELECT * FROM KlausurV2 WHERE Standort='Garching'),
     KlausurH2 as(SELECT * FROM KlausurV2 WHERE Standort<>'Garching')
```

3) Schreiben Sie eine SQL-Abfrage, die die Ursprungsrelation aus den Teilrelationen zusammensetzt.

```
select KlausurV2.*, KlausurV1.Note
from KlausurV1,
  (select * from KlausurH1 union select * from KlausurH2) as KlausurV2
where KlausurV1.MatrNr=KlausurV2.MatrNr
```

Hausaufgabe 4

Zeigen Sie, dass die *write-all / read-any* Methode zur Synchronisation replizierter Daten einen Spezialfall der *Quorum-Consensus*-Methode darstellt.

- Für welche Art von Workloads eignet sich dieses Verfahren besonders gut?
- Wie werden Stimmen zugeordnet um *write-all / read-any* zu simulieren?
- Wie müssen die Quoren Q_w und Q_r vergeben werden?

Dieses Verfahren fordert einen sehr großen Aufwand beim Schreiben, aber nur minimalen Aufwand beim Lesen. Daher eignet es sich besonders gut für Workloads in denen wesentlich mehr Daten gelesen als geschrieben werden. Dies entspricht z.B. analytischen Anfragen.
Siehe Übungsbuch.

Hausaufgabe 5

Um Ausfallsicherheit zu garantieren ist ein Datenwert 'A' auf vier Rechnern verteilt. Jeder Rechner hält dabei eine vollständige Kopie von 'A'. Um Konsistenz zu garantieren wird das Quorum-Consensus-Verfahren eingesetzt. Dabei ist jedem Rechner ein Gewicht $w_i(A)$ wie folgt zugewiesen:

Rechner	Kopie	Gewicht
R_1	A_1	3
R_2	A_2	1
R_3	A_3	2
R_4	A_4	2

Das Lesequorum ist $Q_r(A) = 4$ und das Schreibquorum is $Q_w(A) = 5$.

a) Geben Sie **alle** Lesemöglichkeiten für eine Transaktion auf dem Datum 'A' nach dem Quorum-Consensus-Protokoll an.
A_1,A_2
A_1,A_2,A_3
A_1,A_2,A_3,A_4
A_1,A_2,A_4
A_1,A_3
A_1,A_3,A_4
A_1,A_4
A_2,A_3,A_4
A_3,A_4

b) Geben Sie **alle** Schreibmöglichkeiten für eine Transaktion auf dem Datum 'A' nach dem Quorum-Consensus-Protokoll an.
A_1,A_2,A_3
A_1,A_2,A_3,A_4
A_1,A_2,A_4
A_1,A_3
A_1,A_3,A_4
A_1,A_4
A_2,A_3,A_4

c) Zeigen Sie für dieses Beispiel, dass während eine Transaktion T_1 ein Schreibquorum auf A hält es für andere Transaktionen T_x nicht möglich ist ein Lesequorum für A zu bekommen.
Zum Schreiben muss die Transaktion T_1 mindestens Kopien mit einem Gesamtgewicht von 5 gesperrt haben. Insgesamt haben alle Kopien zusammen das Gewicht 8. Somit können maximal Kopien mit einem Gewicht von zusammen 3 übrig bleiben, womit das Lesequorum von 4 nicht mehr erfüllt werden kann.

Hausaufgabe 6

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`.

```
select supp_nation, cust_nation, l_year, sum(volume) as revenue
from (
   select n1.n_name as supp_nation, n2.n_name as cust_nation,
      extract(year from l_shipdate) as l_year,
      l_extendedprice * (1 - l_discount) as volume
   from supplier, lineitem, orders, customer, nation n1, nation n2
   where
      s_suppkey = l_suppkey and o_orderkey = l_orderkey
      and c_custkey = o_custkey and s_nationkey = n1.n_nationkey
      and c_nationkey = n2.n_nationkey and (
         (n1.n_name = 'UNITED␣STATES' and n2.n_name = 'GERMANY')
          or (n1.n_name = 'GERMANY' and n2.n_name = 'UNITED␣STATES'))
    ) as shipping
group by supp_nation, cust_nation, l_year
order by supp_nation, cust_nation, l_year
```