

Übung zur Vorlesung
Einsatz und Realisierung von Datenbanksystemen im SoSe14

Moritz Kaufmann (moritz.kaufmann@tum.de)
<http://www-db.in.tum.de/teaching/ss14/impldb/>

Blatt Nr. 5

Aufgabe 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

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

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).

Siehe Übungsbuch.

Aufgabe 2

Bei einer abgeleiteten horizontalen Zerlegung kann es vorkommen, dass die erzeugten Fragmente nicht disjunkt sind. Charakterisieren Sie, unter welchen Umständen Disjunktheit gewährleistet ist, und unter welchen Umständen dies nicht der Fall ist. Hinweis: Charakterisieren Sie die Beziehung zwischen der primär zerlegten Relation und der davon abhängig fragmentierten Relation.

Welche Voraussetzungen müssen erfüllt sein, so dass eine abgeleitete Fragmentierung vollständig ist?

Siehe Übungsbuch.

Aufgabe 3

Für die Rekonstruierbarkeit der Originalrelation R aus vertikalen Fragmenten R_1, \dots, 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

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

wobei κ ein Schlüsselkandidat aus R ist.

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

Siehe Übungsbuch.

Gruppenaufgabe

Überlegen Sie sich, welche Tupel bei der Anwendung des bloomfilterbasierten Joins in Abbildung 2 übertragen werden. Markieren Sie insbesondere, welche Tupel übertragen werden, obwohl sie keinen Joinpartner finden (sog. *false drops*). Wie kann die Anzahl dieser *false drops* verringert werden? Welche Eigenschaften sollte die Hashfunktion $h(c)$ die bei dieser Joinbearbeitung verwendet wird erfüllen?

- False Drops sind c7 und c8 die übertragen werden, obwohl sie keinen Joinpartner finden.
- Bei sinnvoller Hashfunktion weniger False Drops je größer der Vektor ist. Gute Hashfunktion!
- Hashfunktion sollte möglichst gleichmäßig verteilen.

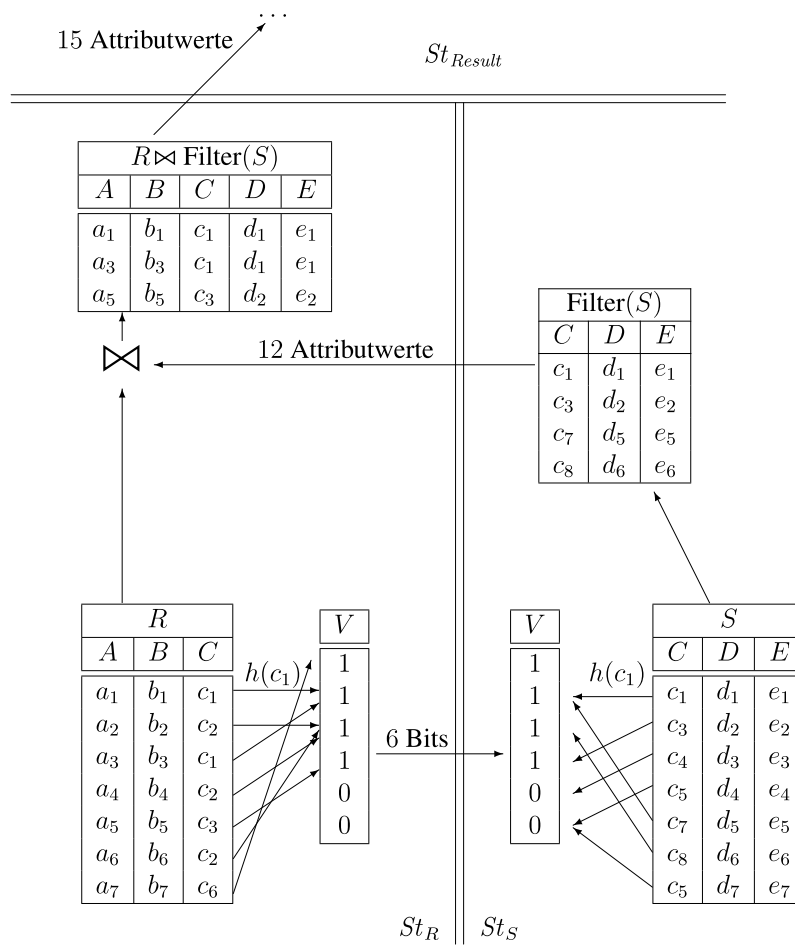


Abbildung 2: Beispiel einer verteilten Joinbearbeitung mit Bloomfilter.