
Aufgabe 1

Zeigen Sie: Wenn R in dritter Normalform ist, dann ist R auch in zweiter Normalform.

Aufgabe 2

Gegeben sei das aus der Vorlesung bekannte Uni-Datenbankschema mit den folgenden Relationen:

- Professoren(PersNr, Name, Rang, Raum)
- Studenten(MatrNr, Name, Semester)
- Vorlesungen(VorlNr, Titel, SWS, gelesenVon)
- Assistenten(PersNr, Name, Fachgebiet, Boss)
- voraussetzten(Vorgänger, Nachfolger)
- hören(MatNr, VorlNr)
- prüfen(MatNr, VorlNr, PersNr, Note)

Welche der folgenden Umsetzungen einer Anfrage in die Algebra läuft (wahrscheinlich) länger? Begründen Sie Ihre Antwort kurz.

Aufgabe 2 a)

1. $\Pi_{Assistenten.name}(\sigma_{Professoren.name='Curie'}(Assistenten \bowtie_{boss=Professoren.persNr} Professoren))$
2. $\Pi_{Assistenten.name}(Assistenten \bowtie_{boss=Professoren.persNr}(\sigma_{Professoren.name='Curie'}(Professoren)))$

Aufgabe 2 b)

1. $\Pi_{name}(\sigma_{semester>5 \wedge Titel='Anfrageoptimierung'} \wedge Studenten.MatrNr=hören.MatrNr (\sigma_{hören.VorlNr=Vorlesungen.VorlNr}(Studenten \times hören \times Vorlesungen)))$
2. $\Pi_{name}(\sigma_{semester>5 \wedge Titel='Anfrageoptimierung'}(Studenten \bowtie hören \bowtie Vorlesungen))$
3. $\Pi_{name}(\sigma_{semester>5}(Studenten) \bowtie hören \bowtie \sigma_{Titel='Anfrageoptimierung'}(Vorlesungen))$

Aufgabe 2 c)

1. $\Pi_{name}(Studenten \bowtie (hoeren \bowtie \sigma_{titel='Anfrageoptimierung'}(Vorlesungen)))$
2. $\Pi_{name}((Studenten \bowtie hoeren) \bowtie \sigma_{titel='Anfrageoptimierung'}(Vorlesungen))$

Aufgabe 3

Gegeben sei folgendes relationales Schema.

- Sachbearbeiter(PersNr, Name, Fachgebiet)
- Kunde(Kundennummer, Name, Adresse)
- betreut(PersNr, Kundennummer)

Ein Benutzer stellt nun die folgende SQL Anfrage.

```
select distinct s.name
from sachbearbeiter s, kunde k, betreut b
where s.PersNr = b.PersNr and b.Kundennummer = k.Kundennummer
and k.name='Schmidt';
```

Aufgabe 3 a)

Welches Ergebnis soll mit der obigen Anfrage vermutlich ermittelt werden?

Aufgabe 3 b)

Überführen Sie die obige SQL Anfrage in einen kanonischen Operatorbaum.

Aufgabe 3 c)

Führen Sie eine (logische) Optimierung des obigen Operatorbaums durch. Verwenden Sie hierfür die in der Vorlesung vorgestellten Heuristiken. Aufgrund fehlender Informationen soll eine Bestimmung der Joinreihenfolge nicht durchgeführt werden.

Aufgabe 4

In einer Anfrage werden die Relationen R , S und T verarbeitet. Die Relationen besitzen die folgenden Schemata.

- $R(\underline{A}, \underline{B}, C)$
- $S(\underline{A}, \underline{D}, E, F)$
- $T(\underline{B}, G, H)$

Aufgabe 4 a)

Betrachten Sie den (optimierten) Operatorbaum der Anfrage in Abbildung 1. Geben Sie zu diesem Operatorbaum den ursprünglichen, kanonischen Operatorbaum an.

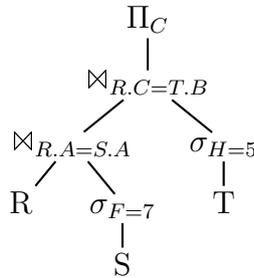


Abbildung 1: Optimierter Operatorbaum

Aufgabe 4 b)

Betrachten Sie den (optimierten) Operatorbaum der Anfrage in Abbildung 2. Geben Sie zu diesem Operatorbaum den ursprünglichen, kanonischen Operatorbaum an.

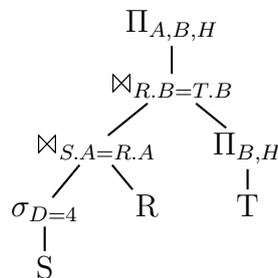


Abbildung 2: Optimierter Operatorbaum

Aufgabe 5

Aufgabe 5 a)

Gegeben sei der algebraische Ausdruck $\sigma_{p_3}(\sigma_{p_1}(R_1) \bowtie_{p_{1,2}} \sigma_{p_2}(R_2))$ und folgende Kardinalitäten (Anzahl der Elemente in den Eingaberelationen) bzw. Selektivitäten (Selektivität der Prädikate):

$$\begin{aligned}
|R_1| &= 460 \\
|R_2| &= 330 \\
\text{Selektivität}(p_1) &= 1/23 \\
\text{Selektivität}(p_2) &= 1/11 \\
\text{Selektivität}(p_3) &= 1/20 \\
\text{Selektivität}(p_{1,2}) &= 1/6
\end{aligned}$$

Geben Sie für jeden Teilausdruck die Ergebniskardinalität an.

Aufgabe 5 b)

Für die Durchführung von Joins wurden in der Vorlesung verschiedene Implementierungsansätze vorgestellt. Diskutieren Sie die Vor- und Nachteile der verschiedenen Join-Implementierungen am Beispiel von Hash-Join, Merge-Join und (blockwise) Nested-Loop Join.

Aufgabe 5 c)

In einem einfachen Datenbanksystem steht nur ein Merge-Join zur Verfügung. Welche Auswirkungen hat dies auf die Plangenerierung? Können in diesem System beliebige Join-Prädikate unterstützt werden?

Aufgabe 6

Der Optimierer eines Datenbanksystems ist noch nicht ganz ausgereift und produziert für eine Anfrage folgenden Operatorbaum:

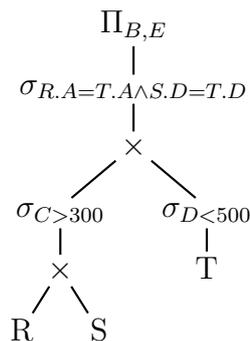


Abbildung 3: Voroptimierter Operatorbaum

Die Relationen besitzen folgendes Schema:

- $R(\underline{A}, B, C)$
- $S(\underline{D}, E, F)$

- $T(A, D)$

Weiterhin kennt das Datenbanksystem folgende Statistiken über die Relationen.

Kardinalitäten		Selektivitäten	
R	100	$\sigma_{C>300}$	$\frac{1}{10}$
S	10.000	$\sigma_{T.D<500}$	$\frac{1}{10}$
T	100.000	$R \bowtie T$	$\frac{1}{1000}$
		$S \bowtie T$	$\frac{1}{1000}$

Aufgabe 6 a)

Geben Sie für obigen Plan für jeden Teilausdruck dessen Kardinalität (d.h. die Anzahl der von diesem Teilausdruck produzierten Tupel) an.

Aufgabe 6 b)

Verbessern Sie den oben abgebildeten Operatorbaum mit Hilfe der in der Vorlesung vorgestellten Heuristiken.

Aufgabe 6 c)

Geben Sie für den in Aufgabenteil b) erstellten, optimierten Operatorbaum für jeden Teilausdruck die Ergebniskardinalität an.

Aufgabe 7

In einer Anfrage werden die Relationen R , S und T verarbeitet. Die Relationen besitzen die folgenden Schemata.

- $R(\underline{RID}, A)$
- $S(\underline{SID}, B)$
- $T(\underline{TID}, SID, C)$

```
select distinct r.a
from R r, S s, T t
where r.rid = s.sid and s.sid = t.sid and s.b = 3 and t.c = 4;
```

Aufgabe 7 a)

Überführen Sie die SQL-Anfrage in einen kanonischen Operatorbaum.

Aufgabe 7 b)

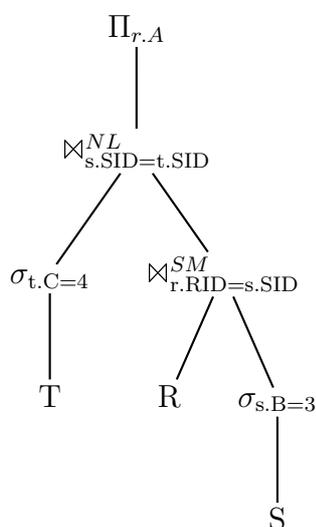
Führen Sie eine (logische) Optimierung des in Aufgabenteil a) erstellten kanonischen

Operatorbaums durch. Verwenden Sie hierfür die in der Vorlesung vorgestellten Heuristiken. Aufgrund fehlender Informationen soll eine Bestimmung der Joinreihenfolge nicht durchgeführt werden.

Aufgabe 7 c)

Zu obiger Anfrage sei der untenstehende Operatorbaum gegeben. Die Superskripte der Join Operatoren geben die jeweils verwendete Joinimplementierung (SM = Sort-Merge Join und NL = Nested-Loop Join) an. Die beiden Eingaben des Sort-Merge Joins sind auf dem jeweiligen Joinattribut sortiert. Gehen sie vom Iteratorenkonzept aus, so dass keine Zwischenergebnisse materialisiert/zwischengespeichert werden! Weiterhin seien die Kardinalitäten und Selektivitäten einer Datenbank aus der folgenden Tabelle gegeben.

Bestim-



Kardinalitäten		Selektivitäten	
R	1.000	s.B = 3	$\frac{1}{5}$
S	500	r.RID = s.SID	$\frac{1}{100}$
T	200	s.SID = t.SID	$\frac{1}{200}$
		t.C = 4	$\frac{1}{2}$

men Sie für die Ausführung des Operatorsbaum auf der Datenbank die Anzahl der Tupel, auf die insgesamt in den Basisrelationen zugegriffen wird. Beachten Sie, dass mehrfache Zugriffe auf ein Tupel mehrfach gezählt werden.

Aufgabe 8

Gegeben sei der unvollständige Auswertungsplan aus untenstehender Abbildung 4. Fügen Sie an den gekennzeichneten Stellen, wenn nötig, Sortieroperationen ein. Geben Sie genau an, nach welchen Attributen sortiert werden muss. Der resultierende Plan soll möglichst wenige Sortieroperationen enthalten.

Weiterhin ist bekannt, dass die TableScans auf Relationen R und T keine sortierte Ausgabe liefern und dass der IndexScan auf einen B^+ -Baum zugreift, dessen Ausgabe sortiert auf den Schlüsseln des Indexes, $S.B$, $S.D$, ist.

Aufgabe 9

Zeichnen Sie einen B-Baum (Grad $k = 1$) der Höhe *eins* mit *maximaler* Speicherauslas-

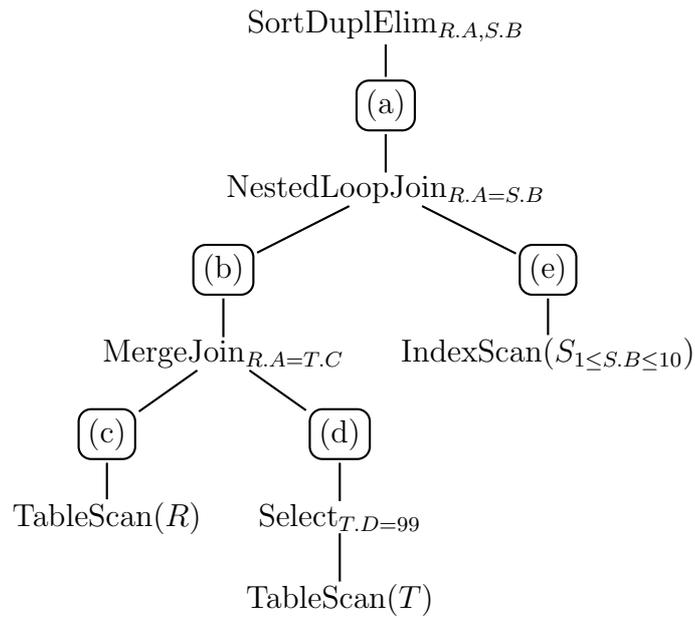


Abbildung 4: ein unvollständiger Auswertungsplan

tung. Tragen Sie nur Schlüssel aus dem Wertebereich von 1 - 20 ein. Bringen Sie die von Ihnen eingetragenen Schlüssel in eine Reihenfolge, so dass der von Ihnen gezeichnete B-Baum entsteht.