Grundlagen von Datenbanken

4. Übung: Algebraische Optimierung

November 2012

Algebraische Optimierung

Ziel

- Effiziente Ausführung eines algebraischen Ausdrucks
- Minimierung der Größe von Zwischenergebnissen
  (das Endergebnis soll gleich bleiben!)

Algebraische Optimierung

Ziel

- Effiziente Ausführung eines algebraischen Ausdrucks
- Minimierung der Größe von Zwischenergebnissen
  (das Endergebnis soll gleich bleiben!)

Voraussetzung

- Abschätzung der Größe von Zwischenergebnissen

Algebraische Optimierung

Ziel

- Effiziente Ausführung eines algebraischen Ausdrucks
- Minimierung der Größe von Zwischenergebnissen
  (das Endergebnis soll gleich bleiben!)

Voraussetzung

- Abschätzung der Größe von Zwischenergebnissen

Verwendete Daten

- Anzahl der Tupel in einer Relation R: $Card(R)$
- Anzahl der unterschiedlichen Werte eines Attributes A_i: j_i
- Vertauschungsregeln für Operationen (siehe Skript)
  z. B.: $\sigma_{P_1}(\sigma_{P_2}(R)) = \sigma_{P_1 \wedge P_2}(R)$

Algebraische Optimierung: Operatorenbaum

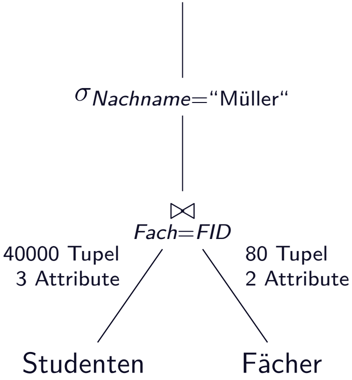
$$\sigma_{Nachname=\text{“Müller“}}(\text{Studenten} \underset{Fach=FID}{\bowtie} \text{Fächer})$$

Algebraische Optimierung: Operatorenbaum

$$\sigma_{Nachname=\text{“Müller“}}(\text{Studenten} \underset{Fach=FID}{\bowtie} \text{Fächer})$$

Annahme für das Beispiel

$Card(\text{Studenten}) = 40000$,
$Card(\text{Fächer}) = 80$
Anzahl unterschiedlicher Namen: 250
(bekannt aus dem Data-Dictionary)



Algebraische Optimierung: Operatorenbaum

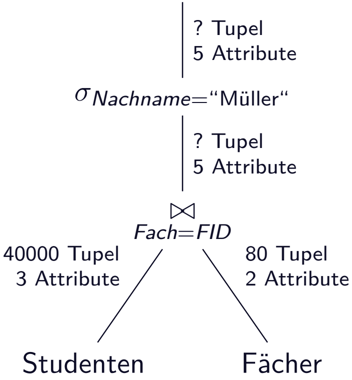
$$\sigma_{Nachname=\text{“Müller“}}(\text{Studenten} \underset{Fach=FID}{\bowtie} \text{Fächer})$$

Annahme für das Beispiel

$Card(\text{Studenten}) = 40000$,
$Card(\text{Fächer}) = 80$
Anzahl unterschiedlicher Namen: 250
(bekannt aus dem Data-Dictionary)

Gesucht

Kardinalitäten beliebiger Operationen,
z. B.: $Card(\sigma_{\ldots}(\text{Studenten} \bowtie \text{Fächer}))$



Selektivitätsfaktor

Motivation

- beschreibt Erwartungswert für die Anzahl der Tupel, die ein Prädikat erfüllen
- basiert auf statistischen Werten
- Annahmen
  - Gleichverteilung der Attributwerte eines Attributes
  - stochastische Unabhängigkeit verschiedener Attribute

Eigenschaften

- $0 \leq \mathrm{SF} \leq 1$
- $Card(\sigma_P(R)) = \mathrm{SF}(P) \cdot Card(R)$

Berechnung des Selektivitätsfaktors

Prädikate bezüglich eines Attributes

- $\mathrm{SF}(A_i = x_i) = \frac{1}{j_i}$, falls Anzahl der Werte j_i für A_i bekannt
- $\mathrm{SF}(A_i \geq x_i \wedge A_i \leq x_j) = \frac{x_j - x_i}{max - min}$, falls bekannt
- ... (siehe Skript)

Zusammengesetzte Prädikate

- $\mathrm{SF}(p(A) \wedge p(B)) = \mathrm{SF}(p(A)) \cdot \mathrm{SF}(p(B))$
- $\mathrm{SF}(p(A) \vee p(B)) = \mathrm{SF}(p(A)) + \mathrm{SF}(p(B)) - (\mathrm{SF}(p(A)) \cdot \mathrm{SF}(p(B)))$
- $\mathrm{SF}(\neg p(A)) = 1 - \mathrm{SF}(p(A))$

Kardinalitätsberechnung beim Verbund

Situation

in der Regel n:1-Verbund zwischen zwei Tabellen:

- TabelleA($\underline{\text{PriA}}$, A_1, ..., A_n, Fremd)
- TabelleB($\underline{\text{PriB}}$, B_1, ..., B_n)
- Referenz: TabelleA.Fremd $\rightarrow$ TabelleB.PriB

Kardinalitätsberechnung beim Verbund

Situation

in der Regel n:1-Verbund zwischen zwei Tabellen:

- TabelleA(PriA, $A_1, \ldots, A_n$, Fremd)
- TabelleB(PriB, $B_1, \ldots, B_n$)
- Referenz: TabelleA.Fremd $\rightarrow$ TabelleB.PriB

Verbund über alle Daten

$$Card(\text{TabelleA} \underset{Fremd=PriB}{\bowtie} \text{TabelleB}) = Card(\text{TabelleA})$$

Kardinalitätsberechnung beim Verbund

Situation

in der Regel n:1-Verbund zwischen zwei Tabellen:

- TabelleA($\underline{\text{PriA}}$, $A_1, \dots, A_n$, $\underline{\text{Fremd}}$)
- TabelleB($\underline{\text{PriB}}$, $B_1, \dots, B_n$)
- Referenz: TabelleA.Fremd $\rightarrow$ TabelleB.PriB

Verbund über alle Daten

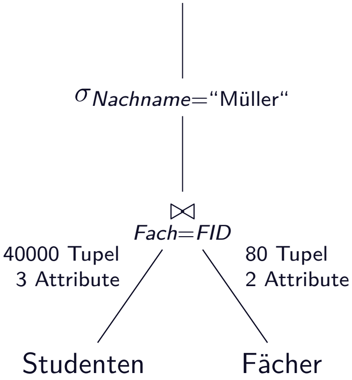
$$Card(\text{TabelleA} \underset{Fremd=PriB}{\bowtie} \text{TabelleB}) = Card(\text{TabelleA})$$

Verbund über eine Teilmenge der Daten

$$Card\left(\sigma_{P_A}(\text{TabelleA}) \underset{Fremd=PriB}{\bowtie} \sigma_{P_B}(\text{TabelleB})\right)$$
$$= \text{SF}(P_A) \cdot \text{SF}(P_B) \cdot Card(\text{TabelleA})$$

Kardinalitätsberechnung im Beispiel

$$\sigma_{Nachname=\text{“Müller“}}(\text{Studenten} \underset{Fach=FID}{\bowtie} \text{Fächer})$$

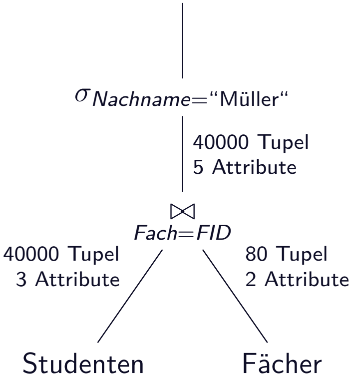


Kardinalitätsberechnung im Beispiel

$$\sigma_{Nachname=\text{“Müller“}}\left(\text{Studenten} \underset{Fach=FID}{\bowtie} \text{Fächer}\right)$$

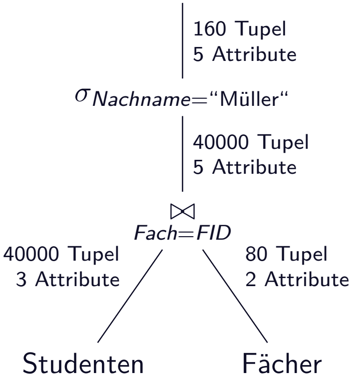
Berechnung des Verbundes

$$Card\left(\text{Studenten} \underset{Fach=FID}{\bowtie} \text{Fächer}\right)$$
$$= Card(\text{Studenten}) = 40000$$



Kardinalitätsberechnung im Beispiel

$$\sigma_{Nachname=\text{“Müller“}}(\text{Studenten} \underset{Fach=FID}{\bowtie} \text{Fächer})$$



Berechnung des Verbundes

$$Card\left(\text{Studenten} \underset{Fach=FID}{\bowtie} \text{Fächer}\right)$$
$$= Card(\text{Studenten}) = 40000$$

Berechnung der Selektion

$$\text{SF}(Nachname = \text{“Müller“}) = \frac{1}{250}$$

Heuristische Regeln zur Optimierung

1. Führe Selektion so früh wie möglich aus
2. Führe Projektion so früh wie möglich aus
3. (Verknüpfe Folgen von unären Operatoren (soweit möglich))
4. Fasse einfache Selektionen auf einer Relation zusammen
5. Verknüpfe bestimmte Selektionen mit einem vorausgehenden Kartesischen Produkt zu einem Verbund
6. (Berechne gemeinsame Teilbäume nur einmal)
7. Bestimme die Verbundreihenfolge so, dass die Anzahl und Größe der Zwischenobjekte minimiert wird
8. Verknüpfe bei Mengenoperationen immer zuerst die kleinsten Relationen

Optimierung des Beispiels

$$\sigma_{Nachname=\text{“Müller“}}(\text{Studenten} \underset{Fach=FID}{\bowtie} \text{Fächer})$$