

Kapitel 6: Anfrageoptimierung basierend auf relationaler Algebra

Klemens Böhm

IWM: Queryoptimierung mit relationaler Algebra – 1

Algebra – Motivation

- Anfragesprache (SQL) ist definitiv nützlich (Datenunabhängigkeit).
- Wir betrachten in diesem Kapitel aber anderen Mechanismus, relationale Algebra:
 - ◆ Einfacher,
 - ◆ erlaubt Aussagen zur Ausführungsreihenfolge.

Einleitung
Projektion
Selektion
Join
Mengen-op.
Vollst.keit

Klemens Böhm

IWM: Queryoptimierung mit relationaler Algebra – 2

Anfragealgebren

- Mathematik – Algebra definiert durch Wertebereich und auf diesem definierte Operatoren.
- Für Datenbankanfragen – Inhalte der Datenbank sind Werte, und Operatoren definieren Funktionen zum Berechnen von Anfrageergebnissen.

Einleitung
Projektion
Selektion
Join
Mengen-op.
Vollst.keit

Klemens Böhm

IWM: Queryoptimierung mit relationaler Algebra – 3

Laufendes Beispiel

Ausleih	Invrnr	Name
	4711	Meyer
	1201	Schulz
	0007	Müller
	4712	Meyer

Einleitung
Projektion
Selektion
Join
Mengen-op.
Vollst.keit

Buch	Invrnr	Titel	ISBN	Autor
	0007	Dr. No	3-125	James Bond
	1201	Objektbanken	3-111	Heuer
	4711	Datenbanken	3-765	Vossen
	4712	Datenbanken	3-891	Ullman
	4717	Pascal	3-999	Wirth

Klemens Böhm

IWM: Queryoptimierung mit relationaler Algebra – 4

Projektion (1)

- Beispiel 1: Projektion auf ein Attribut

$\pi[\text{Name}](\text{Ausleih})$

ergibt als Ergebnisrelation

Name
Meyer
Schulz
Müller

- Doppelte Ergebnistupel eliminiert.

Einleitung
Projektion
Selektion
Join
Mengen-op.
Vollst.keit

Projektion (2)

- Beispiel 2: Projektion auf Attributmenge

$\pi[\text{Invnr}, \text{ISBN}](\text{Buch})$

ergibt

Invnr	ISBN
0007	3-125
1201	3-111
4711	3-765
4712	3-891
4717	3-999

Einleitung
Projektion
Selektion
Join
Mengen-op.
Vollst.keit

Projektion (3)

- Einfache Optimierungsregel:
Bei vielen Projektionen hintereinander
reicht die zuletzt ausgeführte auch allein

$\pi[\text{Invnr}](\pi[\text{Invnr}, \text{ISBN}](\text{Buch}))$

ergibt optimiert

$\pi[\text{Invnr}](\text{Buch})$

- Grafische Darstellung oft hilfreich.
- Wieso sind derartige Optimierungen wichtig?
- Für Anfrageausführung –
Alternativen sind zwar äquivalent,
aber unterschiedlich teuer in der Ausführung.
- Optimierung nutzt Äquivalenz aus,
um günstige Ausführung zu finden.

Einleitung
Projektion
Selektion
Join
Mengen-op.
Vollst.keit

Projektion (4)

- Ersetzung von

$\pi[\text{Invnr}](\pi[\text{Invnr}, \text{ISBN}](\text{Buch}))$

durch

$\pi[\text{Invnr}](\text{Buch})$

ist immer vorteilhaft.

- Bei anderen Transformationen
hängt Vorteilhaftigkeit vom Datenbankzustand ab,
z. B. Vertauschung von Selektionen.

Einleitung
Projektion
Selektion
Join
Mengen-op.
Vollst.keit

Selektion (1)

- Beispiel

$\sigma[\text{Name} \leq 'N'](\text{Ausleih})$

ergibt

Invnr	Name
4711	Meyer
0007	Müller
4712	Meyer

Einleitung
Projektion
Selektion
Join
Mengen-op.
Vollst.keit

Selektion (2)

- Selektionsbedingungen:

- ◆ F *Konstanten-Selektion*

Attribut θ Konstante

boolesches Prädikat θ ist = oder \neq ,
bei linear geordneten Wertebereichen
auch \leq , $<$, \geq oder $>$

- ◆ F *Attribut-Selektion*

Attribut1 θ Attribut2

- ◆ F logische Verknüpfung mehrerer Konstanten- oder Attribut-Selektionen mit \wedge , \vee oder \neg .
(Wird gleich diskutiert.)

Einleitung
Projektion
Selektion
Join
Mengen-op.
Vollst.keit

Selektion (3)

- Einfache Optimierungsregeln:

- ◆ Selektionen lassen sich in der Reihenfolge beliebig vertauschen,

Beispiel: $\sigma_{\text{Invnr}=4711}(\sigma_{\text{Name}\leq'N'}(\text{Ausleih}))$
 $= \sigma_{\text{Name}\leq'N'}(\sigma_{\text{Invnr}=4711}(\text{Ausleih}))$

- ◆ Manchmal lassen sich

Projektion und Selektion vertauschen;

– Ist $\pi_{\text{Invnr}}(\sigma_{\text{Name}\leq'N'}(\text{Ausleih}))$
 $= \sigma_{\text{Name}\leq'N'}(\pi_{\text{Invnr}}(\text{Ausleih}))$?

Einleitung
Projektion
Selektion
Join
Mengen-op.
Vollst.keit

Selektion (4)

- Einfache Optimierungsregeln (Forts.):

- ◆ Manchmal lassen sich

Projektion und Selektion vertauschen (Forts.);

– Voraussetzung für Vertauschbarkeit:
Selektionsattribute
kommen in Projektionsliste vor.

– Beispiel: $\pi_{\text{Name}}(\sigma_{\text{Name}\leq'N'}(\text{Ausleih}))$

– Überprüfung der Vertauschbarkeit
setzt Analyse der Selektionsbedingung
voraus.

Einleitung
Projektion
Selektion
Join
Mengen-op.
Vollst.keit

Einleitung
 Projektion
 Selektion
 Join
 Mengen-op.
 Vollst.keit

Selektion (5)

- Selektionsbedingungen:
 - ◆ *F Konstanten-Selektion*
 $\text{Attribut } \theta \text{ Konstante}$
 - ◆ *F Attribut-Selektion*
 $\text{Attribut1 } \theta \text{ Attribut2}$
 - ◆ *F logische Verknüpfung mehrerer Konstanten- oder Attribut-Selektionen mit \wedge , \vee oder \neg .*
 Jedoch:
 - Analyse aufwendig.
 - Z. T. redundant.
 - Negation problembehaftet
 - Sicherheit der Anfrage?!?

Einleitung
 Projektion
 Selektion
 Join
 Mengen-op.
 Vollst.keit

Verbund

- Syntax des (natürlichen) Verbundes (englisch: natural join)
 $\text{Relation1} \bowtie \text{Relation2}$
- Verbund verknüpft Tabellen über gleichbenannten Spalten bei gleichen Attributwerten.

Einleitung
 Projektion
 Selektion
 Join
 Mengen-op.
 Vollst.keit

Laufendes Beispiel

Ausleih	Invnr	Name
	4711	Meyer
	1201	Schulz
	0007	Müller
	4712	Meyer

Buch	Invnr	Titel	ISBN	Autor
	0007	Dr. No	3-125	James Bond
	1201	Objektbanken	3-111	Heuer
	4711	Datenbanken	3-765	Vossen
	4712	Datenbanken	3-891	Ullman
	4717	Pascal	3-999	Wirth

Einleitung
 Projektion
 Selektion
 Join
 Mengen-op.
 Vollst.keit

Verbund: Beispiel (1)

- $\text{Ausleih} \bowtie \text{Buch}$
 ergibt

Name	Invnr	Titel	ISBN	Autor
Müller	0007	Dr. No	3-125	James Bond
Schulz	1201	Objektbanken	3-111	Heuer
Meyer	4711	Datenbanken	3-765	Vossen
Meyer	4712	Datenbanken	3-891	Ullman

- Nicht ausgeliehenes Pascal-Buch verschwindet: Tupel, die keinen Partner finden (dangling tuples), werden eliminiert.
- In SQL: *outer join*, der „dangling tuples“ übernimmt.

Verbund: Beispiel (2)

- $\pi[\text{Autor}](\text{Buch}) \bowtie \pi[\text{Invnr}](\text{Ausleih})$
entartet zu *kartesischem Produkt*.

Autor	Invnr
James Bond	4711
James Bond	1201
James Bond	0007
James Bond	4712
Heuer	4711
Heuer	1201
Heuer	0007
Heuer	4712
Vossen	4711
...	...

- Kartesisches Produkt – alle Paare, die möglich sind.

Einleitung
Projektion
Selektion
Join
Mengen-op.
Vollst.keit

Eigenschaften Verbund

- Verbund kommutativ: $r_1 \bowtie r_2 = r_2 \bowtie r_1$
- Verbund assoziativ: $(r_1 \bowtie r_2) \bowtie r_3 = r_1 \bowtie (r_2 \bowtie r_3)$
- Daher erlaubt: $\bigotimes_{i=1}^p r_i$
- Beispiel dafür, daß Join-Reihenfolge wichtig:

r_1	<table border="1"><tr><th>B</th></tr><tr><td>b</td></tr></table>	B	b
B			
b			

r_2	<table border="1"><tr><th>B</th><th>C</th></tr><tr><td>a</td><td>c1</td></tr><tr><td>...</td><td>...</td></tr><tr><td>a</td><td>c100000</td></tr></table>	B	C	a	c1	a	c100000
B	C								
a	c1								
...	...								
a	c100000								

r_3	<table border="1"><tr><th>B</th><th>D</th></tr><tr><td>a</td><td>d1</td></tr><tr><td>...</td><td>...</td></tr><tr><td>a</td><td>d100000</td></tr></table>	B	D	a	d1	a	d100000
B	D								
a	d1								
...	...								
a	d100000								

Einleitung
Projektion
Selektion
Join
Mengen-op.
Vollst.keit

Kommentare zu vorangegangener Folie

- Join-Reihenfolge ist wichtig.
- Gute Join-Reihenfolge abhängig von den Daten in den Relationen.
- Gute Join-Reihenfolge schwierig zu erkennen.
- Physische Datenunabhängigkeit.
- Keine Festlegung der Join-Reihenfolge in SQL.

Einleitung
Projektion
Selektion
Join
Mengen-op.
Vollst.keit

Mengenoperationen und Umbenennung (1)

- Mengenoperationen der relationalen Algebra setzen gleiches Schema voraus.

Einleitung
Projektion
Selektion
Join
Mengen-op.
Vollst.keit

Mengenoperationen und Umbenennung (2)

Einleitung
Projektion
Selektion
Join
Mengen-op.
Vollst.keit

Buch1	Autor1	Buch2	Autor2
	James Bond		Witt
	Heuer		Vossen
	Vossen		Silberschatz
	Ullman		Meier
	Wirth		Wirth

- *Umbenennung*

$\beta[\text{neu} \leftarrow \text{alt}](\text{relation})$
(bzw. $\beta_{\text{neu} \leftarrow \text{alt}}(\text{relation})$)

ändert Attributnamen von alt in neu.

- Beispiel: $\beta[\text{Autor1} \leftarrow \text{Autor2}](\text{Buch2})$
- Durch Umbenennung nun Vereinigung, Differenz und Durchschnitt möglich.

Mengenoperationen: Vereinigung

$\text{relation1} \cup \text{relation2}$

Beispiel:

$\text{Buch1} \cup \beta[\text{Autor1} \leftarrow \text{Autor2}](\text{Buch2})$

Einleitung
Projektion
Selektion
Join
Mengen-op.
Vollst.keit

Autor1
James Bond
Heuer
Vossen
Ullman
Wirth
Witt
Silberschatz
Meier

Mengenoperationen: Differenz

$\text{relation1} - \text{relation2}$

Beispiel:

$\text{Buch1} - \beta[\text{Autor1} \leftarrow \text{Autor2}](\text{Buch2})$

Einleitung
Projektion
Selektion
Join
Mengen-op.
Vollst.keit

Autor1
James Bond
Heuer
Ullman

Mengenoperationen: Durchschnitt

$\text{relation1} \cap \text{relation2}$

Beispiel:

$\text{Buch1} \cap \beta[\text{Autor1} \leftarrow \text{Autor2}](\text{Buch2})$

Einleitung
Projektion
Selektion
Join
Mengen-op.
Vollst.keit

Autor1
Vossen
Wirth

Mengenoperationen, Umbenennung (1)

Umbenennung ermöglicht

- Verbunde, wo bisher kartesische Produkte ausgeführt wurden (unterschiedliche Attribute werden gleich benannt),
- kartesische Produkte, wo bisher Verbunde ausgeführt wurden (gleiche Attribute werden unterschiedlich genannt),
- Mengenoperationen.

Einleitung
Projektion
Selektion
Join
Mengen-op.
Vollst.keit

Laufendes Beispiel, leicht modifiziert

- Illustration des ersten Bullets der vorangegangenen Folie:

Ausleih	Invnr	Name
	4711	Meyer
	1201	Schulz
	0007	Müller
	4712	Meyer

Einleitung
Projektion
Selektion
Join
Mengen-op.
Vollst.keit

Buch	I-NR	Titel	ISBN	Autor
	0007	Dr. No	3-125	James Bond
	1201	Objektbanken	3-111	Heuer
	4711	Datenbanken	3-765	Vossen
	4712	Datenbanken	3-891	Ullman
	4717	Pascal	3-999	Wirth

Laufendes Beispiel

- Illustration des zweiten Bullets der vorangegangenen Folie:

Ausleih	Invnr	Name
	4711	Meyer
	1201	Schulz
	0007	Müller
	4712	Meyer

Einleitung
Projektion
Selektion
Join
Mengen-op.
Vollst.keit

Buch	Invnr	Titel	ISBN	Autor
	0007	Dr. No	3-125	James Bond
	1201	Objektbanken	3-111	Heuer
	4711	Datenbanken	3-765	Vossen
	4712	Datenbanken	3-891	Ullman
	4717	Pascal	3-999	Wirth

Mengenoperationen, Umbenennung (2)

- Was ist der Natural Join dieser Relationen ohne Umbenennung?

Buch1	Autor	Buch2	Autor
	James Bond		Witt
	Heuer		Vossen
	Vossen		Silberschatz
	Ullman		Meier
	Wirth		Wirth

Einleitung
Projektion
Selektion
Join
Mengen-op.
Vollst.keit

Mengenoperationen, Umbenennung (3)

Beispiel:

Buch1	Autor	Buch2	Autor
	James Bond		Witt
	Heuer		Vossen
	Vossen		Silberschatz
	Ullman		Meier
	Wirth		Wirth

- Wir wollen Paare bilden (James Bond, Witt), (James Bond, Vossen), ..., (Heuer, Witt), ...
- Vorgehen: Umbenennung und Natural Join.

Einleitung
Projektion
Selektion
Join
Mengen-op.
Vollst.keit

Beispiele für Optimierungsregeln

- Vertauschung Selektion und Projektion.
- Zusammenfassung aufeinanderfolgender Projektionsoperatoren.
- Verbund assoziativ: $(r_1 \bowtie r_2) \bowtie r_3 = r_1 \bowtie (r_2 \bowtie r_3)$

Einleitung
Projektion
Selektion
Join
Mengen-op.
Vollst.keit

Anforderungen

- Anfragen (in Anfragesprache) werden i. Allg. abgebildet auf Folge von Algebra-Operatoren.
- Anforderungen an diese Algebra:
 - ◆ **Optimierbarkeit:** Bestehend aus wenigen Operationen, für die es Optimierungsregeln gibt.
 - ◆ **Effizienz:** Jede Operation ist effizient ausführbar. (Im Relationenmodell hat jede Operation eine Komplexität $\leq O(n^2)$, n Anzahl der Tupel einer Relation.)
 - ◆ **Mengenorientiertheit:** Jede Operation soll auf Mengen von Daten gleichzeitig arbeiten, nicht navigierend nur auf einzelnen Elementen (one-tuple-at-a-time).

Einleitung
Projektion
Selektion
Join
Mengen-op.
Vollst.keit

Unabhängigkeit und Vollständigkeit (1)

- Minimale Relationenalgebra: $\Omega = \pi, \sigma, \bowtie, \beta, \cup$ und $-$
- **Relationale Vollständigkeit:** Jede andere Menge von Operationen, genauso mächtig wie Ω .
- **Strenge relationale Vollständigkeit:** Zu jedem Ausdruck mit Operatoren aus Ω gibt es einen Ausdruck auch mit der anderen Menge von Operationen, also ohne Sprachkonstrukte wie z.B. „“, „while“.

Einleitung
Projektion
Selektion
Join
Mengen-op.
Vollst.keit

Unabhängigkeit und Vollständigkeit (2)

Einleitung
 Projektion
 Selektion
 Join
 Mengen-op.
 Vollst.keit

- Ω ist unabhängig:
Kein Operator kann weggelassen werden, ohne Vollständigkeit zu verlieren.
- Andere unabhängige Menge:
 \bowtie durch \times ersetzen.
- Warum wichtig?
 - ◆ Redundanzfreiheit für formale Überlegungen vorteilhaft.
 - ◆ Minimalität bequemer, wenn es darum geht, Vollständigkeit nachzuweisen.

Division: Beispiel

Einleitung
 Projektion
 Selektion
 Join
 Mengen-op.
 Vollst.keit

r_1	Pilot	Flugzeug	r_2	Flugzeug
	Snoopy	707		707
	Snoopy	727		727
	Snoopy	747		747
	Meyer	707		
	Meyer	727		
	Müller	707		
	Müller	727		
	Müller	747		
	Müller	777		
Lüdenscheid	727			

$r_1 \div r_2$	Pilot		$r_1 \div r_3$	Pilot
	Snoopy			Snoopy
	Müller			Meyer
		Müller		

Division aus Ω herleitbar

Einleitung
 Projektion
 Selektion
 Join
 Mengen-op.
 Vollst.keit

- $r_1(R_1)$ und $r_2(R_2)$ gegeben mit $R_2 \subseteq R_1$, $R' = R_1 - R_2$.
- $R_1 = \{\text{Pilot, Flugzeug}\}$, $R_2 = \{\text{Flugzeug}\}$
- Division von r_1 durch r_2

$$r_1 \div r_2 = \pi_{R'}(r_1) - \pi_{R'}((\pi_{R'}(r_1) \bowtie r_2) - r_1)$$

$$r_1 \div r_2 = \pi_{\text{Pilot}}(r_1) - \pi_{\text{Pilot}}(\pi_{\text{Pilot}}(r_1) \bowtie r_2) - r_1$$

r_1	Pilot	Flugzeug	r_2	Flugzeug
	Snoopy	707		707
	Snoopy	727		727
	Snoopy	747		747
	Meyer	707		
	Meyer	727		
	Müller	707		
	Müller	727		
	Müller	747		
	Müller	777		
	Lüdenscheid	727		

Schlußbemerkungen

Was war Gegenstand dieses Kapitels?

- Grundlage: Relationale Algebra,
- Optimierungs-/Transformationsregeln für Algebraausdrücke – Anfrageoptimierung,
- erhebliche Unterschiede zwischen äquivalenten Ausdrücken,
- Zusammenhang zu physischer Datenunabhängigkeit.

Mögliche Prüfungsfragen, beispielhaft

- Geben Sie die Anforderungen an eine relationale Algebra aus der LV wieder und erklären Sie sie.
- Was versteht man unter Join-Reihenfolge? Erläutern Sie die Wichtigkeit dieses Konzepts.
- Geben Sie ein Beispiel für zwei äquivalente, aber unterschiedliche Algebra-Ausdrücke, die einen Selektions- und einen Join-Operator beinhalten.
(Beliebig variierbar mit anderen Operatoren.)
- Was bedeuten Vollständigkeit und Unabhängigkeit im Kontext der relationalen Algebra?
- Warum enthält Ω nicht \cap ?