



4. Grundlagen des Relationenmodells

Inhalt
Grundkonzepte
Abbildung von ER-Diagrammen
Relationenalgebra
Algebraische Optimierung



Übersicht (1)

Datenstruktur

- Relation (Tabelle)
 - einzige Datenstruktur
 - alle Informationen ausschließlich durch (atomare) Werte dargestellt
 - zeitinvariante Typinformation: Relationenschema

Operatoren auf (mehreren) Relationen

- Vereinigung, Differenz
- Kartesisches Produkt
- Projektion
- Selektion
- zusätzlich: Grundoperationen (Einfügen, Löschen, Ändern)



Übersicht (2)

Beziehungen

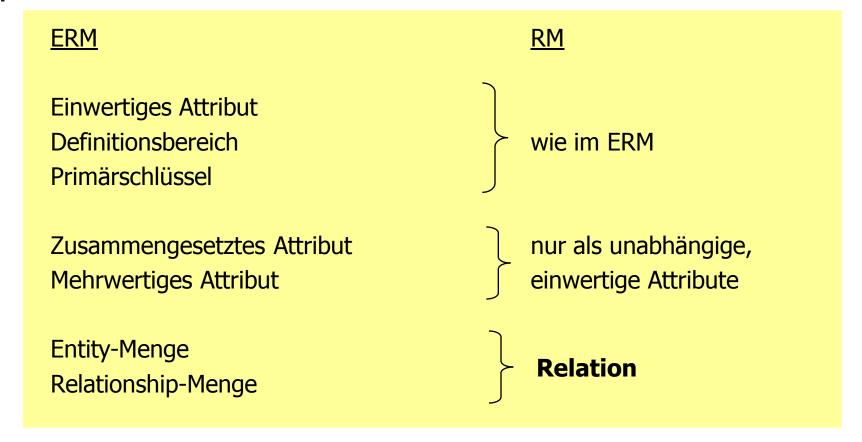
- sind stets explizit, binär und symmetrisch
- werden durch Werte dargestellt
 - Primär-/Fremdschlüssel
 - Gewährleistung von referentieller Integrität
 - können in SQL automatisch gewartet werden (referentielle Aktionen)

Entwurfstheorie

- Normalformenlehre (wünschenswerte und zweckmäßige Relationen)
- Synthese von Relationen



Grundkonzepte (1)



Codd, E.F.: A Relational Model of Data for Large Shared Data Banks, in: Comm. ACM 13:6, June 1970, pp. 377-387.

Grundkonzepte (2)

Mathematische Notation:

D₁, D₂, ..., D_n Definitionsbereiche (Domänen)

 $R \subseteq D_1 \times D_2 \times ... \times D_n$ Relation (Beziehung)

 $t \in R$ Tupel / Record

Notation für Datenbank-Relationen:

 A_1, A_2, \dots, A_n Attribute

 D_1, D_2, \dots, D_n (primitive) Datentypen

 $R \in Rel (A_1:D_1, ..., A_n:D_n)$ Relation über den Attributen $A_1, ..., A_n$

mit den Domänen D₁, ... , D_n

- Relation kann als Tabelle dargestellt werden
- Relation ist eine Menge
 - Garantie der Eindeutigkeit der Zeilen/Tupel durch Primärschlüssel (und ggf. mehrere Schlüsselkandidaten)



Grundkonzepte (3)

Grundregeln:

- Jede Zeile (Tupel) ist eindeutig und beschreibt ein Objekt der Miniwelt
- Die Ordnung der Zeilen ist ohne Bedeutung; durch ihre Reihenfolge wird keine für den Benutzer relevante Information ausgedrückt
- 3. Die Ordnung der Spalten ist ohne Bedeutung, da sie einen eindeutigen Namen (Attributnamen) tragen
- 4. Jeder Datenwert innerhalb einer Relation ist ein atomares Datenelement
- 5. Alle für den Benutzer bedeutungsvollen Informationen sind ausschließlich durch Datenwerte ausgedrückt
- 6. Es existieren ein Primärschlüssel und ggf. weitere Schlüsselkandidaten



Wie wird "relationenübergreifende" Information dargestellt?

Fremdschlüssel, Definition:

Ein Fremdschlüssel bzgl. einer Relation R1 ist ein Attribut oder eine Attributkombination FS einer Relation R2, für das/die zu jedem Zeitpunkt gilt: zu jedem Wert (ungleich Null) von FS muss ein gleicher Wert des Primärschlüssels PS oder eines Schlüsselkandidaten SK in irgendeinem Tupel von Relation R1 vorhanden sein.

Bemerkungen zu Fremdschlüssel:

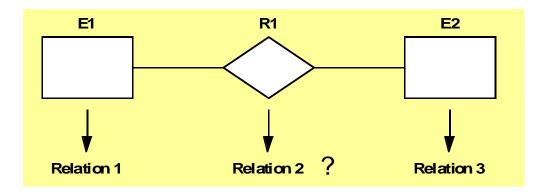
- Fremdschlüssel und zugehöriger Primärschlüssel (Schlüsselkandidat) tragen wichtige inter-relationale (manchmal auch intra-relationale) Informationen. Sie sind auf dem gleichen Wertebereich definiert (vergleichbar und vereinigungsverträglich). Sie gestatten die Verknüpfung von Relationen mit Hilfe von Relationenoperationen.
- Fremdschlüssel können Nullwerte aufweisen, wenn sie nicht Teil eines
 Primärschlüssels sind oder wenn nicht explizit NOT NULL spezifiziert ist.



- Wie wird "relationenübergreifende" Information dargestellt? (Forts.)
 - Bemerkungen zu Fremdschlüssel (Forts.)
 - Schlüsselkandidaten können Nullwerte aufweisen, wenn nicht explizit NOT NULL spezifiziert ist.
 - Ein Fremdschlüssel ist zusammengesetzt, wenn der zugehörige Primärschlüssel (Schlüsselkandidat) zusammengesetzt ist.
 - Eine Relation kann mehrere Fremdschlüssel besitzen, welche die gleiche oder verschiedene Relationen referenzieren.
 - Referenzierte und referenzierende Relation sind nicht notwendigerweise verschieden.
 - Zyklen sind möglich.



Abbildung ERM-RM (1)



Kriterien

- Informationserhaltung, d.h. möglichst genaue Übereinstimmung der Semantik (Übernahme aller spezifizierten Eigenschaften)
- Minimierung der Redundanz
- Minimierung des Verknüpfungsaufwandes
- Natürlichkeit der Abbildung
- Keine Vermischung von Objekten
- Verständlichkeit

Abbildung ERM-RM (2)

2 Entity-Mengen mit (1:n)-Verknüpfung

Regel:

 (1:n)-Beziehungen lassen sich, wenn sie keine eigenen Attribute besitzen, ohne eigene Relation darstellen. Hierzu wird in der abhängigen Relation (mit Beziehungskardinalität 1) der Primärschlüssel der referenzierten Relation als Fremdschlüssel verwendet.

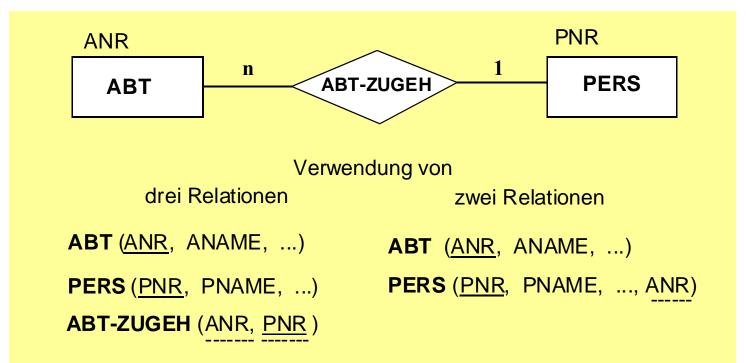


Abbildung ERM-RM (3)

2 Entity-Mengen mit (n:m)-Verknüpfung

Regel:

Eine (n:m)-Relationship-Menge muss durch eine eigene Relation dargestellt werden. Der Primärschlüssel dieser Relation setzt sich aus den Primärschlüsseln der beteiligten Entity-Mengen zusammen. Alle Namen können übernommen werden; es ist jedoch auch eine Umbenennung möglich. Die Attributnamen in einer Relation müssen eindeutig sein.

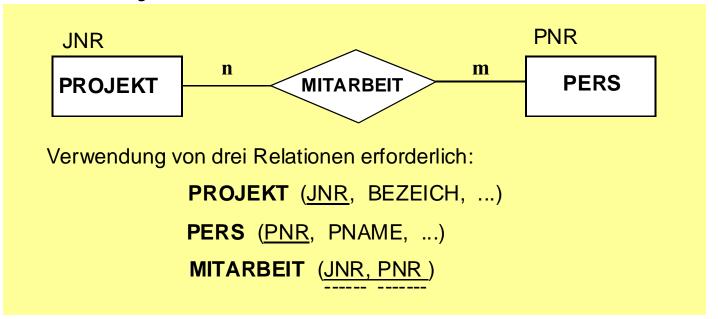


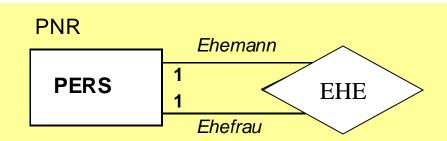


Abbildung ERM-RM (4)

1 Entity-Menge mit (1:1)-Verknüpfung

Regel:

 Der Primärschlüssel der zugehörigen Entity-Menge wird in zwei Rollen verwendet. Deshalb ist eine Umbenennung erforderlich.



1.) Verwendung von zwei Relationen

2.) Verwendung von einer Relation



Abbildung ERM-RM (5)

1 Entity-Menge mit (n:m)-Verknüpfung

Regel:

 Eine (n:m)-Relationship-Menge muss durch eine eigene Relation dargestellt werden. Der Primärschlüssel der zugehörigen Entity-Menge wird in zwei Rollen verwendet. Deshalb ist eine Umbenennung erforderlich.

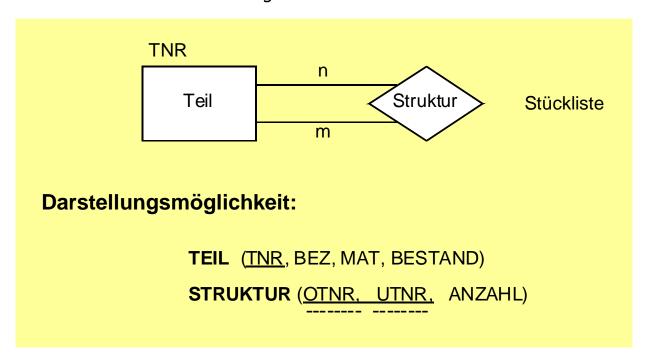


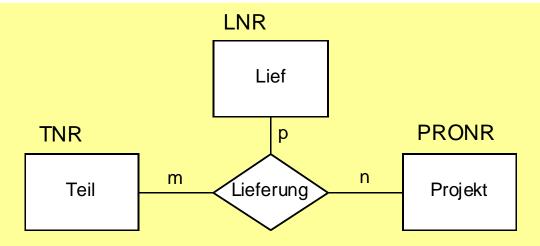
Abbildung ERM-RM (6)

1 Entity-Menge mit (n:m)-Verknüpfung (Forts.)

TEIL	<u>TNR</u>	BEZ	MAT	BESTAND	
	Α	Getriebe	-	10	_
	В	Gehäuse	Alu	0	
	С	Welle	Stahl	100	
	D	Schraube	Stahl	200	
	E	Kugellager	Stahl	50	
	F	Scheibe	Blei	0	
	G	Schraube	Chron	n 100	A
STRU	KTUR	OTNR U A A B B C C	TNR B C D D G D	ANZAHL 1 5 8 4 2 4 2	B 8 C 2 G D E

Abbildung ERM-RM (7)

mehrere (>2) Entity-Mengen mit (n:m)-Verknüpfung



Darstellungsmöglichkeit im RM:

```
LIEF (LNR, LNAME, LORT, ...)

PROJEKT (PRONR, PRONAME, PORT, ...)

TEIL (TNR, TBEZ, GEWICHT, ...)

LIEFERUNG (LNR, PRONR, TNR, ANZAHL, DATUM)
```



Abbildung ERM-RM (8)

mehrere (>2) Entity-Mengen mit (n:m)-Verknüpfung (Forts.)

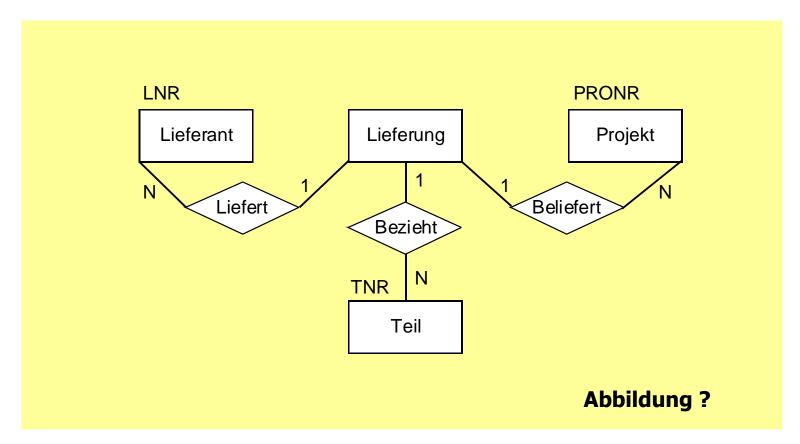
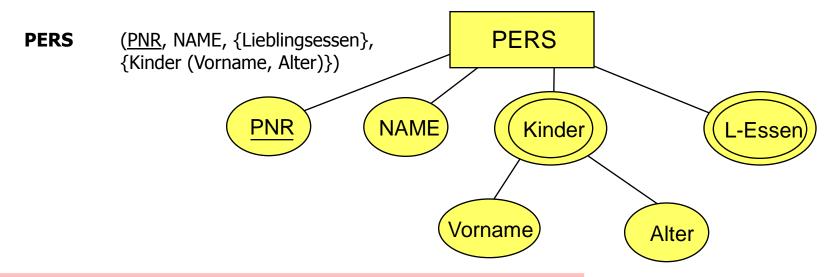




Abbildung ERM-RM (9)

Abbildungstypen innerhalb einer Entity-Menge (Forts.)



Darstellungsmöglichkeit:

PERS (PNR, NAME ...)

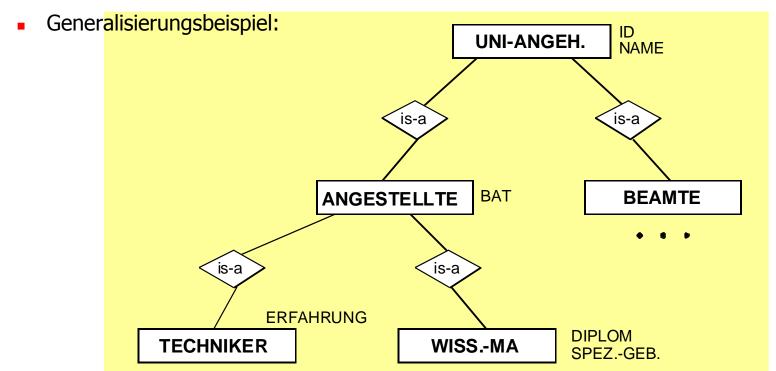
L-ESSEN (PNR, GERICHT)

KINDER (PNR, VORNAME, ALTER)



Abbildung ERM-RM (10)

- Generalisierung
 - RM sieht keine Unterstützung der Abstraktionskonzepte vor
 - keine Maßnahmen zur Vererbung (von Struktur, Integritätsbedingungen, Operationen)
 - "Simulation" der Generalisierung eingeschränkt möglich





- Generalisierung: Hausklassenmodell
 - Jede Instanz ist genau einmal und vollständig in ihrer Hausklasse gespeichert
 - Es wird eine horizontale Partitionierung der DB-Instanzen erreicht

				ID	NAME			
						111	Ernie	
			ANGES	STELLTE	1	I NAME	BAT	
			007				arfield	la
TECHNIKER ID				ERFAHI	RUNG	I NA	ME	BAT
			123	SUN		Do	nald	IVa
WISSMA. ID DIPLOM		SEPZGEB.		' NA	AME	BAT		
	333	Info	rmatik	RECOV	ERY	Da	aisy	lla
	765	Mat	hematik	ERM		Gr	ouch	lla

1

Abbildung ERM-RM (13)

- Generalisierung: Hausklassenmodell (Forts.)
 - Eigenschaften:
 - niedrige Speicherungskosten und keine Änderungsanomalien
 - Retrieval kann rekursives Suchen in Unterklassen erfordern
 - explizite Rekonstruktion durch Relationenoperationen (π, \cup)
 - <u>Beispiel</u>: Finde alle ANGESTELLTE: $\pi_{\text{ID, NAME, BAT}}(\text{TECHNIKER}) \cup \pi_{\text{ID, NAME, BAT}}(\text{WISS.-MA}) \cup \text{ANGESTELLTE}$

Projektion: unärer Operator der Relationenalgebra zur Auswahl von Spalten/Attributen; siehe unten.

Detaillierte Behandlung der relationalen Operatoren weiter hinten in diesem Kapitel!

Abbildung ERM-RM (14)

- Generalisierung: Partitionierungs-Modell
 - Jede Instanz wird entsprechend der Klassenattribute in der Is-a-Hierarchie zerlegt und in Teilen in den zugehörigen Klassen gespeichert
 - Es wird nur das ID-Attribut dupliziert
 - Es wird eine vertikale Partitionierung in der DB erzielt

UNI-ANGEH.	ID	NAME		ANGE	ST	ELLTE	ID	BAT	
	007	Garfield					007	la	
	111	Ernie					123	IVa	
	123	Donald					333	lla	
	333	Daisy					765	lla	
	765	Grouch							
TEC			ECH	HNIKER		ID	ERF	AHRUNG	
						123	S	UN	
	WISSMA			ID	DIPLOM		SPEZGEB		
				333	Informatik		Informatik ERM		
				765	M	athemat	ik N	1AD	



Abbildung ERM-RM (15)

- Generalisierung: Partitionierungs-Modell (Forts.)
 - Eigenschaften
 - geringfügig erhöhte Speicherungskosten, aber hohe Aufsuch- und Aktualisierungskosten
 - Integritätsbedingungen: TECHNIKER.ID ⊆ ANGESTELLTE.ID, usw.
 - Instanzenzugriff erfordert implizite oder explizite Verbundoperationen ()
 - Beispiel: Finde alle TECHNIKER-Daten
 TECHNIKER ANGESTELLTE UNI-ANGEH.
 ID

Join/Verbund: binärer Operator der Relationenalgebra zur Verbindung der Tupeln der Argumentrelationen über gleiche Attributwerte; siehe unten.

Detaillierte Behandlung der relationalen Operatoren weiter hinten in diesem Kapitel!

Abbildung ERM-RM (16)

- Generalisierung: Volle Redundanz
 - Volle Redundanz
 - Eine Instanz wird wiederholt in jeder Klasse, zu der sie gehört, gespeichert
 - Sie besitzt dabei die Werte der Attribute, die sie geerbt hat, zusammen mit den Werten der Attribute der Klasse

UNI-ANGEH.	ID	NAME		ANGESTELLTE	ID	NAME	BAT
	007	Garfield	<u> </u>		007	Garfield	la
	111	Ernie			123	Donald	IVa
	123	Donald			333	Daisy	lla
	333	Daisy			765	Grouch	lla
	765	Grouch					
		l	1			'	

TECHNIKER	ID	NAME	BAT	ERFAHRUNG
	123	Donald	IVa	SUN

WISSMA	ID	NAME	BAT	DIPLOM	SPEZGEB.
		Daisy		imonnatin	RECOVERY
	765	Grouch	lla	Mathematik	ERM



Abbildung ERM-RM (17)

Generalisierung: Volle Redundanz (Forts.)

Eigenschaften

- sehr hoher Speicherplatzbedarf und Auftreten von Änderungsanomalien
- sehr einfaches Retrieval, da nur die Zielklasse (z. B. ANGESTELLTE) aufgesucht werden muss



Abbildung ERM-RM (20)

Zusammenfassung der Abbildungskonzepte

- Datenstruktur: Relation (Tabelle)
 - einzige Datenstruktur (neben atomaren Werten)
 - alle Informationen ausschließlich durch Werte dargestellt
 - Integritätsbedingungen auf/zwischen Relationen: relationale Invarianten

Abbildung von Beziehungen durch PS/SK – FS

- alle Beziehungen sind explizit, binär und symmetrisch
- alle Beziehungstypen müssen im Prinzip durch (n:1)-Beziehungen dargestellt werden
- (n:m)-Beziehungstypen sind durch eine eigene Relation darzustellen
- ein (n:1)-Beziehungstyp wird in der Regel nur dann auf eine eigene Relation abgebildet, wenn er beschreibende Attribute besitzt

Abstraktionskonzepte

- keine direkte Bereitstellung der Abstraktionskonzepte, z.B. Generalisierung
- begrenzte Möglichkeiten zur Abbildung



Relationenalgebra - Operatoren (1)

- Algebra: nicht leere Menge von Objekten + Familie von Operationen
- Operationen
 - Klassische Mengenoperationen:
 - Vereinigung, Differenz, kartesisches Produkt
 - ableitbar: Durchschnitt
 - Relationenoperationen:
 - Projektion, Restriktion (Selektion)
 - ableitbar: Verbund (Join), Division
- Auswahlvermögen entspricht Relationenkalkül ("relational vollständig")



Relationenalgebra - Operatoren (2)

- Selektion (Restriktion): σ_P
 - Auswahl von Zeilen einer Relation über ein Prädikat
 - P = log. Formel (ohne Quantoren!) bestehend aus Attributnamen, Konstanten,
 Vergleichsoperatoren (< , = , > , ≤ , ≠, ≥) und logischen Verknüpfungen
 (∨ , ∧ , ¬)
 - $\sigma_P(R) = \{ t \mid t \in R \land P(t) \}$
 - Beispiel:

ERG :=
$$\sigma_{ANR='K55' \land GEHALT > 50000}$$
 (PERS)

PER	S <u>PNR</u>	NAME	ALTER	GEHALT	ANR	MNR
	406	COY	47	50 700	K55	123
	123	MÜLLER	32	43 500	K51	-
	829	SCHMID	36	45 200	K53	777
	574	ABEL	28	36 000	K55	123

4

Relationenalgebra - Operatoren (3)

- Projektion: π
 - Auswahl von Spalten (Attribute) A₁, A₂, ..., A_k aus einer Relation R (Grad n >= k)
 - $\pi_{A_1, A_2, ..., A_k}$ (R) = { p | ∃ t ∈ R : p = < t [A_1], ..., t [A_k] >} (Alternative: Benutzung von Spaltennummern)
 - Duplikateliminierung
 - Beispiel:

$$\pi_{\text{ANR, MNR}}$$
 (PERS)

PERS	<u>PNR</u>	NAME	ALTER	GEHALT	ANR	MNR
	406	COY	47	50 700	K55	123
	123	MÜLLER	32	43 500	K51	-
	829	SCHMID	36	45 200	K53	777
	574	ABEL	28	36 000	K55	123



Relationenalgebra - Operatoren (4)

- Klassische Mengenoperationen
 - Voraussetzung: Gleicher Grad und Vereinigungsverträglichkeit der beteiligten Relationen
 - Basisoperatoren

Vereinigung:
$$R \cup S = \{t \mid t \in R \lor t \in S\}$$

Differenz:
$$R - S = \{t \mid t \in R \land t \notin S\}$$

Redundante Operatoren

Durchschnitt:
$$R \cap S = R - (R - S) = \{t \mid t \in R \land t \in S\}$$

Symmetrische Differenz:
$$R \triangleright S = (R \cup S) - (R \cap S)$$

Relationenalgebra - Operatoren (5)

- Erweitertes Kartesisches Produkt
 - $K = R \times S = \{ k \mid \exists x \in R, y \in S : (k = x | y) \}$

nicht $<< x_1,, x_r>, < y_1,, y_s>>$	
wie 'übliches' kartesisches Produkt!	

mit $x | y = \langle x_1, ..., x_r, y_1, ..., y_s \rangle$,

PERS	<u>PNR</u>	ALTER	ANR
	406	47	K55
	123	32	K51
	829	36	K53

ABT	<u>ANR</u>	ANAME	ORT
	K51	PLAN.	KL
	K53	EINK.	F

ABT x PERS	ANR	ANAME	ORT	PNR	ALTER	ANR'
	K51	PLAN.	KL	406	47	K55
	K51	PLAN.	KL	123	32	K51
	K51	PLAN.	KL	829	36	K53
	K53	EINK.	F	406	47	K55
	K53	EINK.	F	123	32	K51
	K53	EINK.	F	829	36	K53

4

Relationenalgebra - Operatoren (6)

- Verbund, Join, ⊕-Join
 - Seien R und S Relationen, Θ ∈ {<, =, >, ≤, ≠, ≥} (arithm. Vergleichsoperator), A Attribut von R und B Attribut von S. Θ-Verbund zwischen R und S:

$$V = (R \bowtie_{A \Theta B} S) = \sigma_{A \Theta B} (R \times S)$$

Alternative Definition anhand Spaltennummern
 Annahme: R hat Grad r und S hat Grad s, 1 ≤ i ≤ r, 1 ≤ j ≤ s,

$$V = (R \bowtie_{i \Theta} S) = \sigma_{i \Theta} r + j (R \times S)$$

- Gleichverbund (⊕ = "=")
 - Ein Gleichverbund zwischen R und S heißt verlustfrei, wenn alle Tupel von R und S am Verbund teilnehmen (sonst verlustbehaftet). Die inverse Operation Projektion erzeugt dann wieder R und S (lossless join).



Relationenalgebra - Operatoren (7)

- Verbund, Join, ⊕-Join (Forts.)
 - Definition ,fortsetzbar` auf mehrere Join-Attribute
 - Natürlicher Verbund R S: Gleichverbund über alle übereinstimmenden Attribute und anschließende Projektion, so dass keine Attribute doppelt
 - Verlustfreier Verbund:

$$\pi_{\text{ANR, ANAME, ORT}}$$
 (ABT \bowtie PERS) = ABT $\pi_{\text{PNR, ANR, ALTER}}$ (ABT \bowtie PERS) = PERS;

Al	3T	<u>ANR</u>	ANAME	ORT
		K51	PLAN	KL
		K53	EINK.	F
		K55	VERTR.	F

PERS	<u>PNR</u>	ALTER	ANR	
	406	47	K55	
	123	32	K51	
	829	36	K53	
	574	28	K55	

ABT ⋈ PERS	ANR	ANAME	ORT	PNR	ALTER
	K51	PLAN	KL	123	32
	K53	EINK.	F	829	36
	K55	VERTR.	F	406	47
	K55	VERTR.	F	574	28

4

Relationenalgebra - Operatoren (8)

- Definition Natürlicher Verbund
 - gegeben: $R(A_1, A_2, ..., A_{r-j+1}, ..., A_r)$, $S(B_1, B_2, ..., B_j, ..., B_s)$
 - o.B.d.A. (sonst. Umsortierung): $B_1 = A_{r-j+1}$, $B_2 = A_{r-j+2}$, ..., $B_j = A_r$
 - Natürlicher Verbund zwischen R und S:

$$N = R \bowtie S =$$

$$\pi_{A_{1'}, \dots, A_{r'} B_{j+1'}, \dots, B_{S}} (\sigma_{(R.A_{r-j+1} = S.B_{1}) \land \dots \land (R.A_{r} = S.B_{j})} (R \times S))$$



Relationenalgebra - Operatoren (9)

- Natürlicher Verbund Beispiel
 - Finde alle Angestellten (PNR, ALTER, ANAME), die in einer Abteilung in Frankfurt arbeiten und zwischen 30 und 34 Jahre alt sind.

ABT	<u>ANR</u>	ANAME	AORT
	K51	Planung	Kaiserslautern
	K53	Einkauf	Frankfurt
	K55	Vertrieb	Frankfurt

PERS	<u>PNR</u>	NAME	ALTER	GEHALT	ANR	MNR
	406	Coy	47	50 700	K55	123
	123	Müller	32	43 500	K51	-
	829	Schmid	36	45 200	K53	777
	574	Abel	28	36 000	K55	123



Relationenalgebra - Operatoren (10)

Natürlicher Verbund – Beispiel (Forts.)

Annahmen:

ABT: N/10 Tupel

PERS: N Tupel

Gleichverteilung der Attributwerte

AORT: 20 Werte

Stochastische Unabhängigkeit der Werte verschiedener Attribute

 Verlustfreie Verbunde von R1 und R2 über Primär-/Fremdschlüssel, mit Card(R1) < Card(R2): Card(R1 ⋈ R2) = Card(R2)

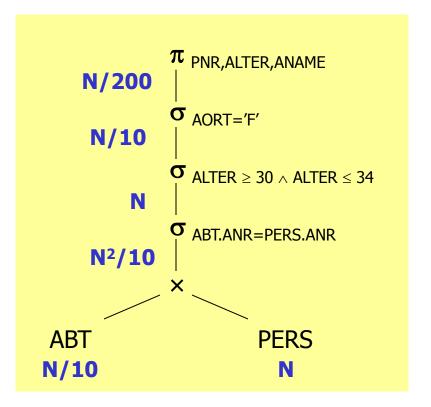


Relationenalgebra - Operatoren (11)

- Natürlicher Verbund Beispiel (Forts.)
 - Lösung 1:

```
\pi_{\text{PNR,ALTER,ANAME}} (\sigma_{\text{AORT}='\text{F}'} (\sigma_{\text{ALTER} \ge 30 \land \text{ALTER} \le 34} (\sigma_{\text{ABT.ANR}=\text{PERS.ANR}} (ABT × PERS))))
```

zugehöriger Operatorbaum:



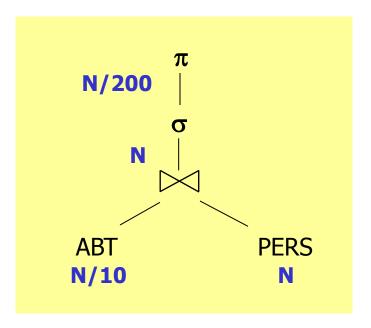


Relationenalgebra - Operatoren (12)

- Natürlicher Verbund Beispiel (Forts.)
 - Lösung 2:

```
\pi_{\text{PNR,ALTER,ANAME}} (\sigma_{\text{ALTER} \ge 30 \land \text{ALTER} \le 34 \land \text{AORT}='F'} (ABT PERS))
```

zugehöriger Operatorbaum:



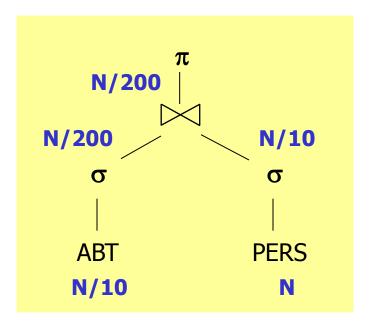


Relationenalgebra - Operatoren (13)

- Natürlicher Verbund Beispiel (Forts.)
 - Lösung 3:

```
\pi_{\text{PNR,ALTER,ANAME}} ((\sigma_{\text{AORT}='F'} (ABT)) (\sigma_{\text{ALTER} \ge 30 \land \text{ALTER} \le 34} (PERS)))
```

zugehöriger Operatorbaum:



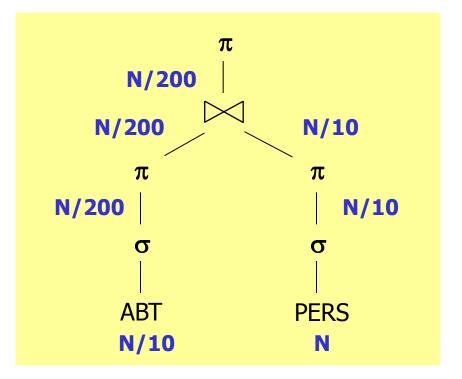


Relationenalgebra - Operatoren (14)

- Natürlicher Verbund Beispiel (Forts.)
 - Lösung 4:

```
\pi_{\text{PNR,ALTER,ANAME}} ((\pi_{\text{ANR,ANAME}} (\sigma_{\text{AORT}='F'}(\text{ABT})))) \qquad (\pi_{\text{PNR,ALTER,ANR}} (\sigma_{\text{ALTER} \ge 30 \, \land \, \text{ALTER} \le 34} (\text{PERS}))))
```

zugehöriger Operatorbaum:





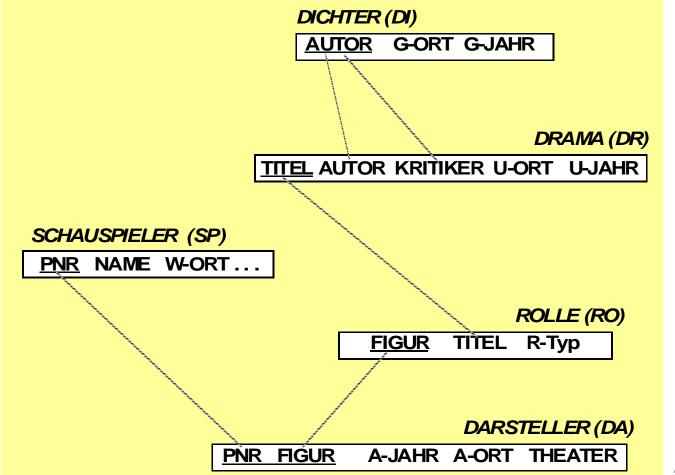
Relationenalgebra – Beispiel (1)

Beispiel - ER-Diagramm: **DICHTER Ist_Autor Ist_Kritiker** _von _von **DRAMA** n **SCHAUSPIELER** hat n Stellt_dar **ROLLE** m



Relationenalgebra – Beispiel (16)

Beispiel – relationales DB-Schema:





Relationenalgebra – Beispiel (17)

- Beispiel Anfragen:
 - Finde alle Schauspieler (NAME), die die Figur "Faust" gespielt haben.

$$π$$
 NAME ($σ$ FIGUR="FAUST" (SP \bowtie DA))

Finde alle Schauspieler (NAME), die im Drama "Faust" mitgespielt haben.

$$\pi_{\text{NAME}}$$
 ($\sigma_{\text{TITEL}} = \text{"Faust"}$ (SP \bowtie DA \bowtie RO))

• Finde alle Schauspieler (NAME), die in Dramen von Schiller mitgespielt haben.

$$\pi_{\text{NAME}}$$
 ($\sigma_{\text{AUTOR}} = \text{"Schiller"}$ (SP \bowtie DA \bowtie RO \bowtie DR))



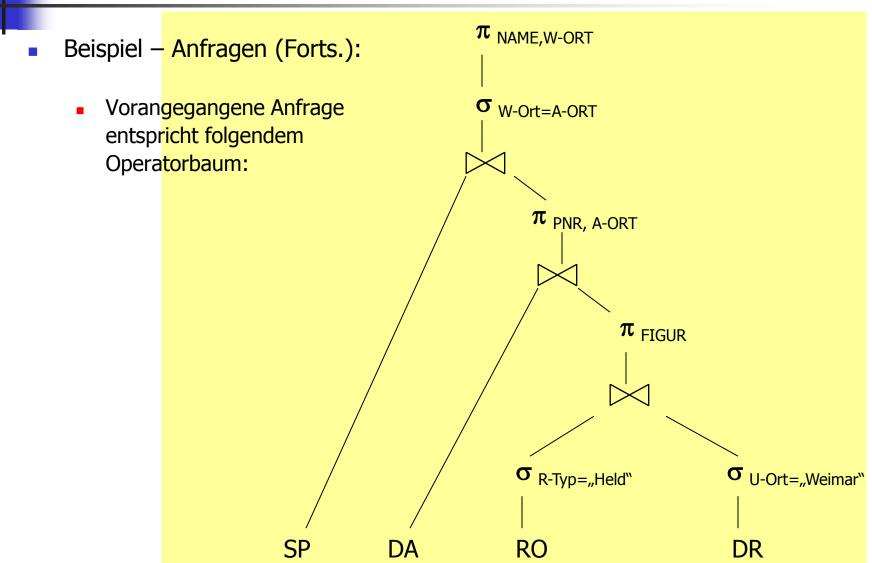
Relationenalgebra – Beispiel (18)

- Beispiel Anfragen (Forts.):
 - Finde alle Schauspieler (NAME, W-ORT), die bei in Weimar uraufgeführten Dramen an ihrem Wohnort als 'Held' mitgespielt haben.

```
\pi_{NAME, W-ORT} (\sigma_{W-ORT= A-ORT} (SP ) (
\sigma_{PNR} (\sigma_{PNR, A-ORT} (DA ) (\sigma_{FIGUR} (
FIGUR

(\sigma_{R-TYP= 'HELD'} RO) ) (\sigma_{U-ORT= `WEIMAR'} DR) ) ) ) ) ) ) ) ) TITEL
```

Relationenalgebra – Beispiel (19)





Relationenalgebra – Beispiel (20)

- Beispiel Anfragen (Forts.):
 - Liste alle Dramen mit ihren Autoren (TITEL, AUTOR, G-JAHR) auf, die nach 1800 uraufgeführt wurden.

$$π$$
 TITEL, AUTOR, G-JAHR ($σ$ U-JAHR > 1800 (DI \bowtie DR))
AUTOR

 Finde alle Schauspieler (NAME, W-ORT), die in Dramen von Schiller, die von in Weimar geborenen Dichtern kritisiert wurden, mitgespielt haben.

```
\pi_{\text{NAME, W-ORT}} (SP \bowtie DA \bowtie RO \bowtie PNR FIGUR TITEL (\sigma_{\text{AUTOR}=,Schiller}" (DR)) \bowtie (\sigma_{\text{G-ORT}=,Weimar}" (DI))) KRITIKER=AUTOR
```



Relationenalgebra – Beispiel (21)

- Beispiel Anfragen (Forts.):
 - Finde die Schauspieler, die nie gespielt haben.

$$\pi_{PNR}$$
 (SP) - π_{PNR} (DA))

Finde die Schauspieler, die nur Faust oder Wallenstein gespielt haben.

$$\pi_{PNR}$$
 (DA) - π_{PNR} ($\sigma_{FIGUR \neq \text{"FAUST"} \land FIGUR \neq \text{"Wallenstein"}}$ (DA)))

 Anfragen wie "Welcher Dichter ist Schauspieler?" oder "Welcher Dichter hat in einem seiner Stücke gespielt?" können "eigentlich" nicht beantwortet werden, da es keine systemkontrollierte Beziehung zwischen **Dichter** und **Schauspieler** gibt.



- ACHTUNG: Connection Trap!
 - Verbund kann im Allg. <u>nicht</u> als Umkehroperation zur Projektion angesehen werden
 - Beispiel: DA1 und DA2 als Projektionen auf DA; DA3 als Verbund von DA1 und DA2

DA1	PNR	A-ORT
	P1	MA
	P1	KL
	P2	MA

DA	PNR	FIGUR	A-ORT	
	P1 P1 P2	Faust Mephisto Wallenstein	MA KL MA	

DA2	DA2 FIGUR		
	Faust	MA	
	Mephisto	KL	
	Wallenstein	MA	

DA3	PNR	FIGUR	A-ORT	
	P1	Faust	MA	
	P1	Wallenstein	MA	
	P1	Mephisto	KL	
	P2	Faust	MA	
	P2	Wallenstein	MA	



Algebraische Optimierung (1)

 Relationenalgebraische Formulierungen spezifizieren Ausführungsreihenfolge (prozedurale Elemente), äquivalente Umformungen möglich

Optimierungsproblem

- gegeben: Ausdruck der Relationenalgebra (RA)
- gesucht: äquivalenter, möglichst effizient auszuführender RA-Ausdruck
- Bestimmung einer möglichst guten Ausführungsreihenfolge (Einsatz von Heuristiken)
- Statistische Kenngrößen werden dem DB-Katalog entnommen
 - $N_i = Card(R_i)$
 - j_i = Anzahl der verschiedenen Werte eines Attributs A_i

Algebraische Optimierung (2)

Rewrite-Regeln

- Kommutativgesetz f
 ür Produkte und Verbunde
 - $R1 \times R2 \equiv R2 \times R1$
 - R1 \bowtie R2 \equiv R2 \bowtie R1
- Assoziativgesetz für Produkte und Verbunde
 - $(R1 \times R2) \times R3 \equiv R1 \times (R2 \times R3)$
 - $(R1 \bowtie R2) \bowtie R3 \equiv R1 \bowtie (R2 \bowtie R3)$
- Zusammenfassung von Folgen von Projektionen
 - $\bullet \quad \pi_{A,B,C} (\pi_{A,B,C,...,Z} (SP)) \equiv \pi_{A,B,C} (SP)$
- Zusammenfassung von Folgen von Selektionen

•
$$\sigma_{F1} (\sigma_{F2} (R)) \equiv \sigma_{F1 \wedge F2} (R) \equiv \sigma_{F2 \wedge F1} (R) \equiv \sigma_{F2} (\sigma_{F1} (R))$$

Steht hier für beliebige ⊕-Verbunde

Algebraische Optimierung (3)

- Rewrite-Regeln (Forts.)
 - Vertauschung von Selektionen und Projektionen
 - F enthält nur Attribute aus A, ..., Z:

$$\sigma_{F}(\pi_{A,...,Z}(R)) \equiv \pi_{A,...,Z}(\sigma_{F}(R))$$

F enthält Attribute aus A, ..., Z, B1, ..., Bm:

$$\pi_{A, ..., Z} (\sigma_{F}(R)) \equiv \pi_{A, ..., Z} (\sigma_{F}(\pi_{A, ..., Z, B1, ..., Bm}(R)))$$

- Vertauschung von Selektion und Kartesischem Produkt
 - F enthält nur Attribute aus R1:

$$\sigma_F(R1 \times R2) \equiv \sigma_F(R1) \times R2$$

• allgemeiner: $F = F1 \land F2 \land F3$

F1 nur auf R1, F2 nur auf R2, F3 auf beiden

$$\sigma_{F}(R1 \times R2) \equiv \sigma_{F1}(R1) \bowtie_{F3} \sigma_{F2}(R2)$$



Algebraische Optimierung (4)

- Annahmen
 - Gleichverteilung der Attributwerte eines Attributes
 - Stochastische Unabhängigkeit der Werte verschiedener Attribute
- Selektivitätsfaktor (SF)
 - basiert auf statistischen Werten
 - beschreibt hinsichtlich eines Qualifikationsprädikats den erwarteten Anteil an Tupeln, die das Prädikat erfüllen
 - 0 ≤ SF ≤ 1
 - Card $(\sigma_p(R)) = SF(p) \cdot Card(R)$



Algebraische Optimierung (5)

SF-Berechnung

j_i:Anzahl der Werte des Attributs A_i

$$A_i = a_i \qquad SF = \begin{cases} 1/j_i & \text{falls } j_i \text{ bekannt, } z.B. \text{ Index auf } A_i \\ 1/10 & \text{sonst} \end{cases}$$

$$A_i = A_k \qquad SF = \begin{cases} 1/max(j_i, j_k) & \text{falls } j_i \text{ und } j_k \text{ bekannt} \\ 1/j_i & \text{falls nur } j_i \text{ bekannt} \\ 1/10 & \text{sonst} \end{cases}$$

$$A_i \geq a \qquad (\text{oder } A_i > a) \qquad SF = \begin{cases} (\text{high-key} - a_i)/(\text{high-key} - \text{low-key}) & \text{falls (keys)} \\ 1/3 & \text{sonst} & (\text{Wert) interpolierbar} \end{cases}$$

$$A_i \geq a_i \wedge A_i \leq a_k \qquad SF = \begin{cases} (a_k - a_i)/(\text{high-key} - \text{low-key}) & \text{falls bekannt} \\ 1/4 & \text{sonst} \end{cases}$$

Algebraische Optimierung (6)

SF-Berechnung bei Ausdrücken

- SF $(p(A) \land p(B)) = SF(p(A)) \cdot SF(p(B))$
- SF $(p(A) \lor p(B)) = SF (p(A)) + SF (p(B)) SF (p(A)) \cdot SF (p(B))$
- SF $(\neg p(A)) = 1 SF(p(A))$

Join-Selektivitätsfaktor (JSF)

- Card (RS) = JSF * Card(R) * Card(S)
- bei (N:1)-Joins (verlustfrei): Card (RS) = Max(Card(R), Card(S))



Algebraische Optimierung (7)

Beispiel

DB-Schema

```
ABT (<u>ANR</u>, BUDGET, A-ORT)
PERS (<u>PNR</u>, NAME, BERUF, GEHALT, ALTER, ANR)
PM (<u>PNR,JNR</u>, DAUER, ANTEIL)
PROJ (<u>JNR</u>, BEZEICHNUNG, SUMME, P-ORT)
```

 Anfrage: Finde Name und Beruf von Angestellten, deren Abteilung in KL ist und die in KL Projekte durchführen.

Annahmen:

ABT: N/5 Tupel

PERS: N Tupel

PM: 5N Tupel

PROJ: M Tupel

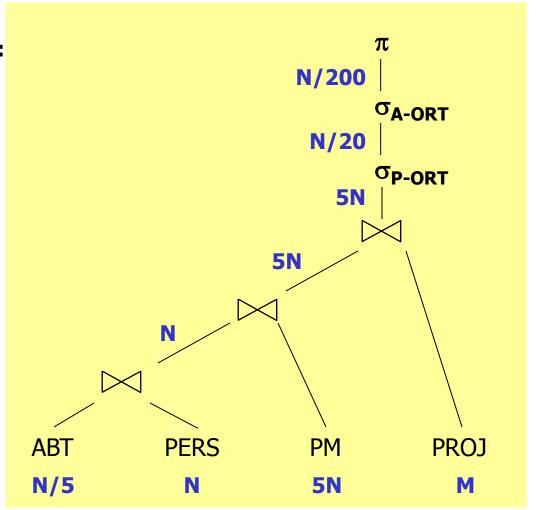
Anzahl der Attributwerte von A-ORT: 10, P-ORT: 100

Verlustfreie Verbunde von R1 und R2 über Primär-/Fremdschlüssel



Algebraische Optimierung (8)

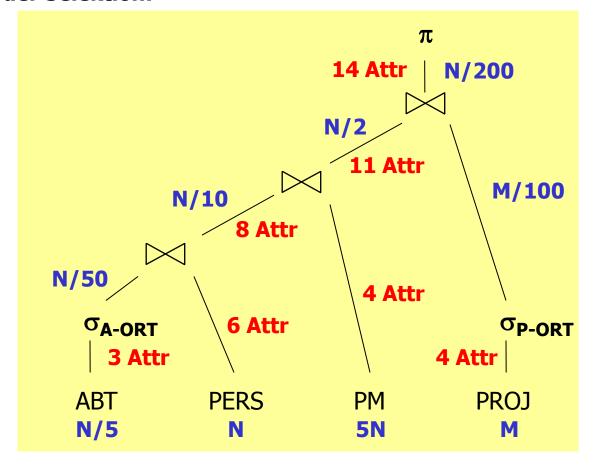
- Beispiel (Forts.)
 - Ausgangslösung:





Algebraische Optimierung (9)

- Beispiel (Forts.)
 - Verschieben der Selektion:





Algebraische Optimierung (9)

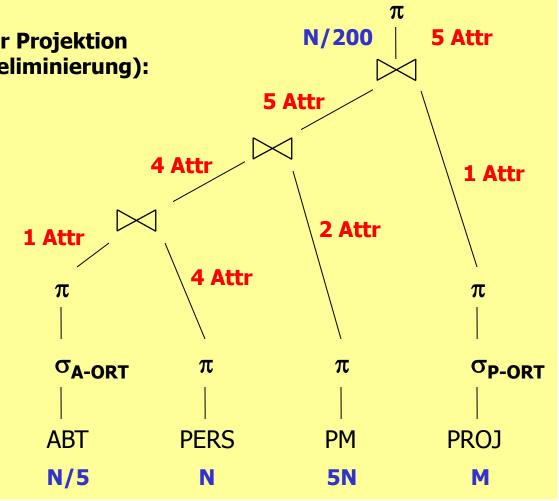
- Beispiel (Forts.)
 - Verschieben der Selektion:

I.
Führe Selektion so früh
wie möglich aus!

Algebraische Optimierung (10)

Beispiel (Forts.)

 Verschieben der Projektion (ohne Duplikateliminierung):



2 Attr



Algebraische Optimierung (10)

- Beispiel (Forts.)
 - Verschieben der Projektion (ohne Duplikateliminierung):

II.

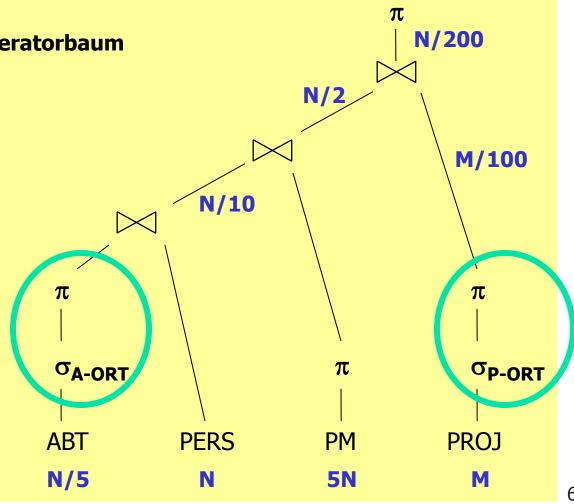
Führe Projektion
(ohne Duplikateliminierung)
so früh wie möglich aus!

Bem.: Der Nutzen einer frühzeitigen Projektionsausführung hängt von mehreren Faktoren ab.

Algebraische Optimierung (11)

Beispiel (Forts.)

Optimierter Operatorbaum (Vorschlag):





Algebraische Optimierung (11)

- Beispiel (Forts.)
 - Optimierter Operatorbaum (Vorschlag):

III.

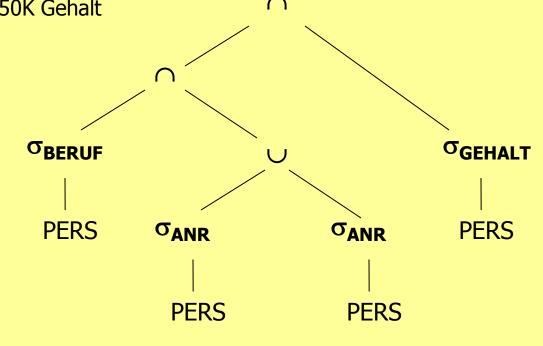
Verknüpfe Folgen von unären
Operatoren wie Selektion und
Projektion (wenn diese tupelweise
abgewickelt werden können)!

Algebraische Optimierung (12)

Weitere Optimierungsmaßnahmen

Ausdrucksauswertung

 Beispiel: Finde alle Programmierer aus Abteilung K51 oder K55 mit mehr als 50K Gehalt



Algebraische Optimierung (12)

- Weitere Optimierungsmaßnahmen (Forts.)
 - Ausdrucksauswertung
 - Beispiel: Finde alle Programmierer aus Abteilung K51 oder K55 mit mehr als 50K Gehalt

```
<sup>™</sup>BERUF="P" ∧ GEHALT>50K
 ∧ (ANR="K51" ∨ ANR="K55")
 |
PERS
```



Algebraische Optimierung (12)

- Weitere Optimierungsmaßnahmen (Forts.)
 - Ausdrucksauswertung
 - Beispiel: Finde alle Programmierer aus Abteilung K51 oder K55 mit mehr als 50K Gehalt

IV.

Fasse einfache Selektionen auf einer Relation zusammen!



Algebraische Optimierung (13)

Weitere Optimierungsmaßnahmen (Forts.)

V.

Verknüpfe bestimmte Selektionen
mit einem vorausgehenden
Kartesischen Produkt
zu einem Verbund!

VI.

Berechne gemeinsame Teilbäume nur einmal (wenn die Zwischenspeicherung der Ergebnisse nicht zu teuer ist)!



Algebraische Optimierung (14)

Weitere Optimierungsmaßnahmen (Forts.)

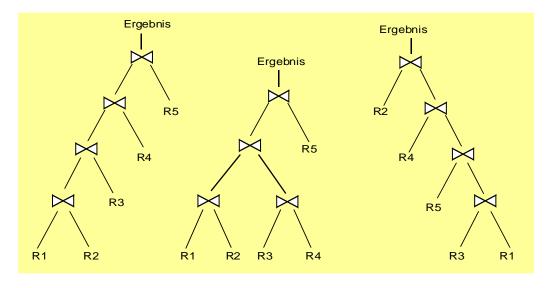
- Kombination von Verbundoperationen
 - Assoziativität und Kommutativität von Verbundoperationen (gilt auch für Vereinigung und Durchschnitt)

Allgemeines Problem bei binären Relationenoperationen

- Was ist die beste Verknüpfungsreihenfolge?
- Im allgemeinen Fall sind n! Reihenfolgen möglich
- Die genaue Größe einer Zwischenrelation ergibt sich erst nach Ende der erzeugenden Operation
 - Dynamische Entscheidung aufwendiger, aber genauer als Abschätzung
 - Bei jedem Auswertungsschritt werden die momentan kleinsten (Zwischen-)Relationen ausgewählt



- Weitere Optimierungsmaßnahmen (Forts.)
 - Kombination von Verbundoperationen (Forts.)
 - Allgemeines Problem bei binären Relationenoperationen (Forts.)
 - Einige Verknüpfungsreihenfolgen für den Verbund mit n=5



VII. Bestimme die Verbundreihenfolge so, dass die Anzahl und Größe der Zwischenobjekte minimiert wird!



Algebraische Optimierung (17)

Weitere Optimierungsmaßnahmen (Forts.)

- Reihenfolge von Mengenoperationen
- Kardinalität der Vereinigung: max(N(R1), N(R2))

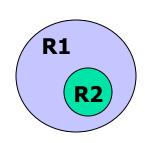
$$\leq$$
 N(R1) + N(R2)

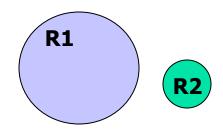


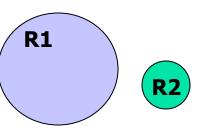
0

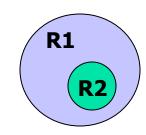
$$\leq$$
 N(R1 \cap R2)

 $\leq \min(N(R1), N(R2))$









VIII. Verknüpfe bei Mengenoperationen immer zuerst die kleinsten Relationen!



Algebraische Optimierung (18)

Heuristische Regeln:

- Führe Selektion so früh wie möglich aus
- Führe Projektion (ohne Duplikateliminierung) frühzeitig aus
- Verknüpfe Folgen von unären Operationen wie Selektion und Projektion
- Fasse einfache Selektionen auf einer Relation zusammen
- Verknüpfe bestimmte Selektionen mit einem vorausgehenden Kartesischen Produkt zu einem Verbund
- Berechne gemeinsame Teilbäume nur einmal
- Bestimme Verbundreihenfolge so, dass die Anzahl und Größe der Zwischenobjekte minimiert wird
- Verknüpfe bei Mengenoperationen immer zuerst die kleinsten Relationen

Weitere Operatoren (1)

- Division
 - Ziel
 - Beantwortung von Fragen, bei denen eine "ganze Relation" zur Qualifikation herangezogen wird
 - Simulation des Allquantors ⇒ ein Tupel aus R steht mit <u>allen</u>
 Tupeln aus S in einer bestimmten Beziehung
 - Definition
 - Sei R vom Grad r und S vom Grad s, r > s und s ≠ 0;
 t sei (r-s)-Tupel, u sei s-Tupel;
 S-Attribute ⊂ R-Attribute;
 - Dann gilt: $\mathbf{R} \div \mathbf{S} = \{\mathbf{t} \mid \forall \mathbf{u} \in \mathbf{S} : (\mathbf{t} | \mathbf{u} \in \mathbf{R})\}$

Weitere Operatoren (2)

- Division (Forts.)
 - Beispiel:

PNR	FIGUR	A-Jahr
P1	Faust	1999
P1	Nathan	1998
P2	Werther	1997
P3	Faust	1998
P3	Nathan	1999
P3	Werther	1998
	P1 P1 P2 P3 P3	P1 Faust P1 Nathan P2 Werther P3 Faust P3 Nathan

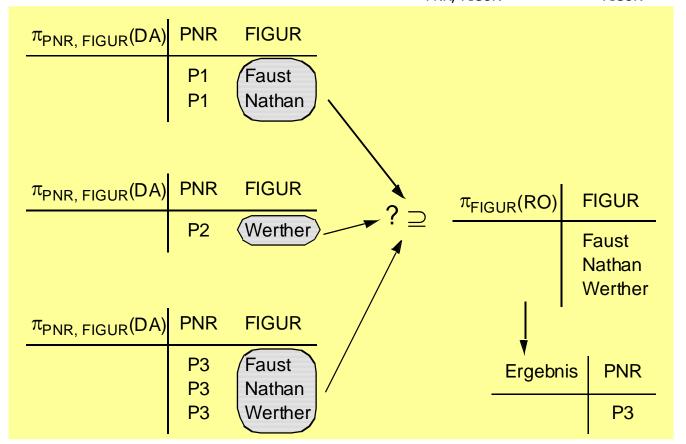
Faust Faust Nathan Nathan der Weise Werther Die Leiden	RO	FIGUR	TITEL	R-Typ
Nathan Nathan der Weise Werther Die Leiden		Faust	Faust	
Werther Die Leiden		Nathan	Nathan der Weis	se
		Werther	Die Leiden	

Welche Schauspieler haben alle Rollen gespielt:

$$(\pi_{PNR, FIGUR}(DA)) \div (\pi_{FIGUR}(RO))$$

Weitere Operatoren (3)

- Division (Forts.)
 - Beispiel (Forts.)
 - Welche Schauspieler haben alle Rollen gespielt: $(\pi_{PNR, FIGUR}(DA)) \div (\pi_{FIGUR}(RO))$



Weitere Operatoren (4)

- Division (Forts.)
 - Beschreibung der Division mit den Grundoperatoren

T =
$$\pi_{1, 2, ..., r-s}$$
 (R)
W = (T × S) - R
V = $\pi_{1, 2, ..., r-s}$ (W)
R ÷ S = T - V
= $\pi_{1, 2, ..., r-s}$ (R) - $\pi_{1, 2, ..., r-s}$ (($\pi_{1, 2, ..., r-s}$ (R) × S) - R)

• Es gilt: $(R \times S) \div S = R$

Weitere Operatoren (5)

- Division (Forts.)
 - Weitere Beispiele
 - Finde alle Schauspieler (NAME), die alle Rollen in Dramen von Goethe gespielt haben.

```
\pi_{\text{Name}}(\text{ (SP}\bowtie(\pi_{\text{PNR,FIGUR}}(\text{DA}))) \div (\pi_{\text{FIGUR}}(\text{RO}\bowtie(\sigma_{\text{AUTOR}=\text{"Goethe}``}(\text{DR}))))))
```

 Finde alle Schauspieler (NAME), die alle Narrenrollen am Pfalztheater gespielt haben.

$$\pi_{\text{Name}}(\text{ (SP}\bowtie(\pi_{\text{PNR,FIGUR}}(\sigma_{\text{THEATER}=...}(\text{DA}))) \div (\pi_{\text{FIGUR}}(\sigma_{\text{R-Typ}=...}(\text{RO}))))$$

Weitere Operatoren (6)

- Intervallverbund (Band Join)
 - Anstatt des arithmetischen Vergleichsoperators Θ des Θ -Joins wird hier eine Intervall-Bedingung überprüft.
 - Grob: Kartesisches Produkt zwischen zwei Relationen R (Grad r) und S (Grad s) eingeschränkt durch eine Intervall-Bedingung zwischen i-Spalte von R und j-Spalte von S.
 - Intervall $I = [c_1, c_2]$ mit c_1, c_2 positive Konstanten, wobei eine größer Null sein muss.
 - Intervall-Verbund zwischen R und S:

$$V = R \bowtie_{i|j} S = \sigma_{i|j} (R \times S) = \sigma_{R.i-c_1 \leq S.j \leq R.i+c_2} (R \times S)$$



Weitere Operatoren (7)

- Intervallverbund (Band Join)
 - Bemerkung
 - Ein Tupel s aus S 'kombiniert' mit einem Tupel r aus R nur, wenn der Wert der j-Spalte von S im Intervall der Größe c_1+c_2 um den Wert der i-Spalte von R liegt.
 - Beispiel: G = σ PNR≠PNR' (PERS ⋈ PERS')
 ALTER [2,2] ALTER'

PERS	PNR	ALTER
	P1	25
	P2	23
	P3	28
		l

G	PNR	ALTER	PNR'	ALTER'	
	P1	25	P2	23	
	P2	23	P1	25	
	l				l



Weitere Operatoren (8)

- Äußerer Verbund (*Outer Join*)
 - Ziel: ,Verlustfreiheit` soll erzwungen werden!
 - Trick: Einfügen spezieller Leerzeilen zur künstlichen Erzeugung von Verbundpartnern
 - Beispiel

SP	PNR	NAME
	P1	X
	P2	у

DA	PNR	FIGUR
	P1	F
	P1	W
	P3	М

Weitere Operatoren (9)

- Äußerer Verbund (Forts.)
 - Definition
 - Seien A die Verbundattribute, "≡" der undefinierte Wert und

$$R' = R \cup ((\pi_A(S) - \pi_A(R)) \times X \times X = ...)$$

$$S' = S \cup ((\pi_A(R) - \pi_A(S)) \times X \times X = ...)$$

Äußerer Gleichverbund:

$$R \longrightarrow S := R' \longrightarrow S'$$

 $R.A=S.A$ $R'.A=S'.A$

Äußerer natürlicher Verbund:

$$R \supset S := R' \supset S'$$

Weitere Operatoren (10)

- Außerer Verbund (Forts.)
 - Linker äußerer Gleichverbund
 - Bei bei dieser Operation bleibt die linke Argumentrelation verlustfrei, d. h.,
 bei Bedarf wird ein Tupel durch "NULL"-Werte "nach rechts" aufgefüllt

$$R \longrightarrow S := R \longrightarrow S'$$

 $R.A=S.A$ $R.A=S'.A$

- Rechter äußerer Gleichverbund
 - Dabei bleibt analog die rechte Argumentrelation verlustfrei; fehlende Partnertupel werden durch Auffüllen mit "NULL"-Werten "nach links" ergänzt

$$R \searrow S := R' \searrow S$$
 $R.A=S.A$
 $R'.A=S.A$

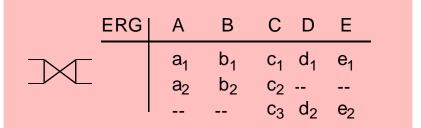


Weitere Operatoren (11)

- Äußerer Verbund (Forts.)
 - Beispiele:

R	Α	В	С
	a ₁	b ₁	c ₁
	a_2	b_2	c_2

	ERG	Α	В	C D	E
$\supset \!$		a ₁	b ₁	c ₁ d ₁	e ₁
		a_2	b ₂	c ₁ d ₁ c ₂	





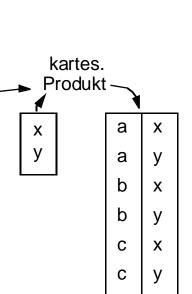
а

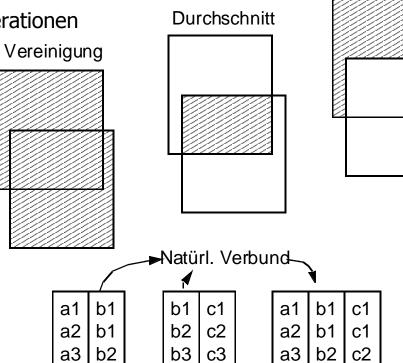
b

C

Relationenalgebra

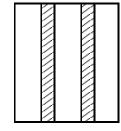
- Zusammenfassung
 - Algebra mit Auswahlvermögen der Prädikatenlogik 1. Stufe
 - Abgeschlossenheit bzgl. der Algebraoperationen
 - Klassische Mengenoperationen
 - Relationenoperationen

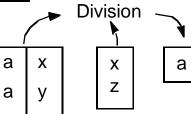




Restriktion

Projektion





a z b x c y

Differenz

83