Das Relationenmodell

Relationale Datenbankverwaltungssysteme

E.F. Codd: 1970

Datenbankverwaltungssysteme:

- INFORMIX
- INGRES
- ORACLE
- PROGRESS
- SYBASE
- SYSTEM R (IBM) -> DB/2
- UNIFY

Überblick

1. Relationenschema, Relation,
   Attribute, Tupel, Schlüssel
2. Operationen auf Relationen:
3. NF^2-Relationen

1. Darstellung v. Objekten u. Beziehungen

in Tabellenform:

Objektklasse / Assoziation ... Tabelle

Attribute .... Spalten in der Tabelle (Attributnamen)

Objektinstanzen bzw. Verknüpfungen .... Zeilen in der Tabelle

Beispiel:

flüge (Flugnr	von	nach	Abflug	Ankunft)
83	JFK	O'Hare	11:30a	1:43p
84	O'Hare	JFK	3:00p	5:55p
109	JFK	L.A.	9:50p	2:52a
213	JFK	Boston	11:43a	12:45p
214	Boston	JFK	2:20p	3:12p

(Tabelle 1)

2. Relationenschema

Ein *Relationenschema R* ist eine endliche Menge von Attributnamen $\{A_1, A_2, \ldots A_n\}$.

Attributnamen werden auch kurz als *Attribute* bezeichnet.

Zu jedem Attribut A_i gibt es eine Menge D_i ($1 < i < n$), den *Wertebereich (domain)* von A_i.

Notation: $Dom(A_i)$ .... Wertebereich von A_i.

Die Wertebereiche sind willkürliche, nichtleere, endliche oder abzählbar endliche Mengen.

Relation

Sei $D = D_1 \cup D_2 \cup D_3 \cup \ldots \cup D_n$.

Eine Relation r auf einem Relationenschema R, r(R), ist eine endliche Menge von Abbildungen $\{t_1, \ldots, t_n)$ von R nach D, wobei für jede Abbildung $t \in r$, $t(A_i)$ aus D_i, $1 \leq i \leq n$, stammt.

Die Abbildungen werden *Tupel* genannt.

Beispiel: Relation

In Tabelle 1 ist das Relationenschema flüge = {Flugnr, von, nach, Abflug, Ankunft}

Dom(Flugnr) = Menge der ein, zwei oder dreistelligen Dezimalzahlen

Dom(von)=Dom(nach)={JFK, O'Hare, L.A., Boston, Atlanta)

Dom(Abflug) = Dom(Ankunft) = Menge der Zeiten eines Tages

Die Tabelle 1 hat 5 Tupel. Eines von ihnen ist t mit t(Flugnr) = 84, t(von) = O'Hare, t(nach) = JFK, t(Abflug) = 3.00p, t(Ankunft) = 5:55p.

A-Wert eines Tupel

Der Wert eines Tupels t für ein Attribut A, t(A) = a, heißt A-Wert von t.

Sei X eine Teilmenge von R. Unter der Einengung der Abbildung t auf die Attributmenge X, t(X), versteht man denjenigen X-Wert, der jedem A ∈ X genau dieselben Werte aus Dom(A) zuordnet wie die Abbildung t.

Beispiel: A-Wert eines Tupel

Sei t das Tupel mit t(flugnr) =84, t(von) = O’Hare, t(nach) = JFK, t(Abflug) = 3:00p, t(Ankunft) = 5:55p.

Der von-Wert von t ist t(von) = O’Hare.

Der {von,nach}-Wert von t ist das Tupel t’ definiert als t’(von) = O’Hare, t’(nach) = JFK.

Definition: Schlüssel

Ein *Schlüssel (key)* einer Relation r(R) ist eine Teilmenge K von R, sodaß für je zwei verschiedene Tupeln t1 und t2 aus r, t1(K) <> t2(K), und keine echte Teilmenge K’ von K diese Eigenschaft hat.

K ist ein *Oberschlüssel (superkey)*, wenn K einen Schlüssel enthält.

Wichtig: Eine Relation kann mehrere Schlüssel besitzen (siehe Kapitel Normalformen).

Beispiel:

{Flugnr} und {von,nach} erfüllen die Schlüsselbedingung in

flüge

(Flugnr	von	nach	Abflug	Ankunft)
83	JFK	O’Hare	11:30a	1:43p
84	O’Hare	JFK	3:00p	5:55p
109	JFK	L.A.	9:50p	2:52a
213	JFK	Boston	11:43a	12:45p
214	Boston	JFK	2:20p	3:12p

Aber: Schlüssel zeitinvariant !

Daher ist nur {Flugnr} Schlüssel.

Primärschlüssel

Schlüssel, die explizit zu einem Relationenschema angeführt sind, heißen ausgezeichnete Schlüssel (designated keys).

Im allgemeinen wird ein Schlüssel als *Primärschlüssel* ausgezeichnet.

Dieser wird im Relationenschema durch Unterstreichen gekennzeichnet.

Beispiel:

flüge (<u>Flugnr</u>,von,nach, Ankunft,Abflug)

3. Operationen auf Relationen

Relationale Algebra (Codd 1972)

3.1 Mengenoperationen

3.2 Selektion

3.3 Projektion

3.4 Verbund, kartesisches Produkt

3.5 Division

Im folgenden bezeichnen r,s Relationen über die Schemata R,S. Indizierte A sind Attribute von R, indizierte B, Attribute von S.

3.1 Mengenoperationen

- Vereinigung
- Durchschnitt
- Differenz von Relationen,

die über der gleichen Attributmenge definiert sind.

Beispiel:

r

A	B	C
a1	b1	c1
a1	b2	c1
a2	b1	c2

s

A	B	C
a1	b2	c1
a2	b2	c1
a2	b2	c2

$r \cap s$

A	B	C
a1	b2	c1

$r \cup s$

A	B	C
a1	b1	c1
a1	b2	c1
a2	b1	c2
a2	b2	c1
a2	b2	c2

$r - s$

A	B	C
a1	b1	c1
a2	b1	c2

3.2 Selektion

Auswahl von Zeilen

$\sigma_{A=a}(r) = \{ t \in r \mid t(A) = a \}$

Der **Selektion**soperator ist kommutativ bezüglich der Zusammensetzung:

$\sigma_{A1=a}(\sigma_{A2=b}(r)) = \sigma_{A2=b}(\sigma_{A1=a}(r))$

Beispiel: Selektion

flüge

(Flugnr	von	nach	Abflug	Ankunft)
83	JFK	O'Hare	11:30a	1:43p
84	O'Hare	JFK	3:00p	5:55p
109	JFK	L.A.	9:50p	2:52a
213	JFK	Boston	11:43a	12:45p
214	Boston	JFK	2:20p	3:12p

$\sigma_{VON = \text{“JFK”}}(\text{flüge}) =$

(Flugnr	von	nach	Abflug	Ankunft)
83	JFK	O'Hare	11:30a	1:43p
109	JFK	L.A.	9:50p	2:52a
213	JFK	Boston	11:43a	12:45p

Erweiterung der Selektion

Sei Θ eine Menge von Vergleichsoperatoren, und θ ein Operator aus dieser Menge:

$\sigma_{A\theta a}(r) = \{ t \in r \mid t(A)\ \theta\ a \}$

$\sigma_{A1\theta A2}(r) = \{ t \in r \mid t(A1)\ \theta\ t(A2) \}$

Beispiel: Selektion

flugzeiten

(Flugnr	Abflug	Ankunft)
84	3:00p	5:55p
109	9:40p	2:42a
117	10:05p	10:43a
213	11:43a	12:45p
214	2:20p	3:12p

$\sigma_{\mathbf{ANKUNFT} \leq \mathbf{1.00p}}$ (flugzeiten) =

(Flugnr	Abflug	Ankunft)
109	9:40p	2:42a
117	10:05p	10:43a
213	11:43a	12:45p

3.3 Projektion

Auswahl von Spalten

Sei $X \subseteq R$:

$\pi_{X(r)} = \{ t(X) \mid t \in r \}$

Der **Projektion**soperator ist kommutativ bezüglich der Selektion, wenn die Attribute, auf denen selektiert wird, in den projizierten Attributen enthalten sind:

$$\pi_{A1,A2} (\sigma_{A1 = 5}(r)) = \sigma_{A1 = 5} (\pi_{A1,A2}(r))$$

Beispiel: Projektion

flugzeiten

(Flugnr	Abflug	Ankunft)
84	3:00p	5:55p
109	9:40p	2:42a
117	10:05p	12:43a
213	11:43a	12:45p
214	2:20p	3:12p

$\pi_{ABFLUG,ANKUNFT}$ (flugzeiten) =

(Abflug	Ankunft)
3:00p	5:55p
9:40p	2:42a
10:05p	12:43a
11:43a	12:45p
2:20p	3:12p

3.4 Verbund (Join)

- Definition Verbund
- Kartesisches Produkt
- Verlustfreie Zerlegbarkeit
- Eigenschaften des Verbunds
- Gleichverbund
- Theta-Verbund

Definition Verbund

Der **Verbundoperator** verknüpft zwei Relationen auf ihren gemeinsamen Attributen.

$$r \bowtie s = \{ t \mid \exists t_r \in r \text{ und } \exists t_s \in s : t_r = t(R) \text{ und } t_s = t(S) \}$$

Der Verbundoperator ist kommutativ.

Auch bekannt als *natürlicher Verbund* (*natural Join*)

Beispiel: Verbund

benützbar

(FLUG	MASCHINE)
83	727
83	747
84	727
84	747
109	707

zugelassen

(PILOT	MASCHINE)
Simmons	707
Simmons	727
Barth	747
Hill	727
Hill	747

optionen = benützbar $\bowtie$ zugelassen

(FLUG	MASCHINE	PILOT)
83	727	Simmons
83	727	Hill
83	747	Barth
83	747	Hill
84	727	Simmons
84	727	Hill
84	747	Barth
84	747	Hill
109	707	Simmons

Kartesisches Produkt

Beispiel:

r

(A	B)
a1	b1
a2	b1

s

(C	D)
c1	d1
c2	d1
c2	d2

$r \times s$

(A	B	C	D)
a1	b1	c1	d1
a1	b1	c2	d1
a1	b1	c2	d2
a2	b1	c1	d1
a2	b1	c2	d1
a2	b1	c2	d2

Annahme: $R \cap S = \{\ \}$.

Achtung: Falls $R \cap S = \{\ \}$, dann ist r $\bowtie$ s gleich dem kartesischen Produkt.

Verlustfreie Zerlegbarkeit (1)

R = AB, S = BC

(a) verlustfrei zerlegbar:

q

(A	B	C)
a	b	c
a	b’	c’

r

(A	B)
a	b
a	b’

s

(B	C)
b	c
b’	c’

r $\bowtie$ s = q’

(A	B	C)
a	b	c
a	b’	c’

Verlustfreie Zerlegbarkeit (2)

(b) q ist nicht verlustfrei in r,s zerlegbar:

q’ = r $\bowtie$ s, enthält 2 überschüssige Tupel

q

(A	B	C)
a	b	c
a	b’	c’
a’	b	c’

r

(A	B)
a	b
a	b’
a’	b

s

(B	C)
b	c
b’	c’
b	c’

r $\bowtie$ s = q’

(A	B	C)
a	b	c
a	b’	c’
a	b	c’
a’	b	c’
a’	b	c

Weitere Eigenschaften Verbundoperator

Sei $q = s \bowtie r$, $r' = \pi_R(q)$, dann gilt:

$r' \subseteq r$.

Beispiel:

r

A	B
a	b
a	b'

s

B	C
b	c

$r \bowtie s = q$

A	B	C
a	b	c

$\pi_{AB}(q) =$ r'

A	B
a	b

Weitere Eigenschaften Verbundoperator

Sei q eine Relation über RS,

$q' = \pi_R(q) \bowtie \pi_S(q)$, so gilt $q' \supseteq q$.

Wenn $q = q'$ gilt, dann ist das Relationenschema RS verlustfrei zerlegbar (decomposes losslessly) in R und S.

Gleichverbund (Equijoin)

Seien r (R), s(S) Relationen,

$A_i \in R$, $B_i \in S$, und $Dom(A_i) = Dom(B_i)$, $1 \leq i \leq n$.

Der **Gleichverbund** von r und s über die Attribute $A_1, ..., A_n$ und $B_1, ..., B_n$, geschrieben als

$$r\, [\, A_1 = B_1, ..., A_n = B_n \,]\, s,$$

ist die Relation:

$$q(RS) = \{t \mid \exists\, t_s \in s \land \exists\, t_r \in r \text{ mit } t_r(R) = t,\ t_s(S) = t$$
$$\text{und } t(A_i) = t(B_i),\ 1 \leq i \leq n\ \}$$

Beispiel:

route

FLUG	VON	NACH
84	O'Hare	JFK
109	JFK	L.A.
117	Atlanta	Boston
213	JFK	Boston
214	Boston	JFK

stationiert

PILOT	FLUGHAFEN
Terhune	JFK
Temple	Atlanta
Taylor	Atlanta
Tarbell	Boston
Todd	L.A.
Truman	O'Hare

route [VON = FLUGHAFEN] stationiert = kannfliegen

FLUG	VON	NACH	PILOT	FLUGHAFEN
84	O'Hare	JFK	Truman	O'Hare
109	JFK	L.A.	Terhune	JFK
117	Atlanta	Boston	Temple	Atlanta
117	Atlanta	Boston	Taylor	Atlanta
213	JFK	Boston	Terhune	JFK
214	Boston	JFK	Tarbell	Boston

Theta-Verbund (Theta-join)

Seien r(R), s(S) Relationen, wobei $R \cap S = \{\}$.

Sei $A \in R$, $B \in S$, und die Elemente aus Dom(A) sind mit den Elementen aus Dom(B) mittels der Operatoren aus θ vergleichbar.

Sei $\theta \in \Theta$.

Der Thetaverbund r [A θ B] s ist die Relation:

q(RS) =

$\{ t \mid \exists t_s \in s \wedge \exists t_r \in r \text{ mit } t(R) = t_r, t(S) = t_s \text{ und } t_r(A)\ \theta\ t_s(B) \}$

Beispiel

flugzeitenab

(FLUGNR	ABFLUG	ANKUNFT
60	9:40a	11:45a
91	12:50p	2:47p
112	4:05p	6:15p
306	8:30p	10:25p
420	9:15p	11:11p

flugzeitenbc

(FLUGNR	ABFLUG	ANKUNFT)
11	8:30a	9:52a
60	12:25p	1:43p
156	4:20p	5:40p
158	7:10p	8:35p

flugzeitenab [ANKUNFT < ABFLUG] flugzeitenbc = verbindungen

(FLUGNR	ABFLUG	ANKUNFT	(FLUGNR'	ABFLUG'	ANKUNFT')
60	9:40a	11:45a	60	12:25p	1:43p
60	9:40a	11:45a	156	4:20p	5:40p
60	9:40a	11:45a	158	7:10p	8:35p
91	12:50p	2:47p	156	4:20p	5:40p
91	12:50p	2:47p	158	7:10p	8:35p
112	4:05p	6:15p	158	7:10p	8:35p

3.5 Division

Seien r(R) und s(S) Relationen mit $S \subseteq R$. $R' = R - S$,

dann ist $r \div s$ die Relation:

$$r'(R') = \{ t \mid \forall t_s \in s\ \exists t_r \in r ((t_r(R') = t) \wedge (t_r(S) = t_s)) \}$$

fliegt

(Pilot	Maschine)
Maier	727
Maier	747
Huber	767
Müller	747

q

(Maschine)
727
747

fliegt $\div$ q

(Pilot)
Maier

4. NF^2-RELATIONEN

NF^2: Non First Normal Form Relations (Relational Model with Relation-Valued Attributes)

- **Erweiterung des Relationalen Modells**
- **Darstellung komplexer Objekte und hierarchischer Strukturen (z.B.: CAD, Büroumgebung,...)**
- **Attributwerte sind Relationen**

Schek, Pistor 1982

AUSDRÜCKE DER RELATIONEN-ALGEBRA (RA-QUERIES)

buch

(INVNR	*AUTOR*	*TITEL*	*VERLAG)*
1	**Date**	**Intro DBS I**	**AW**
2	**Date**	**Intro DBS II**	**AW**
3	**Ullman**	**Princ. DBS**	**CSP**
4	**Kroenke**	**DB Proc.**	**SRA**

leser

(LNR	*NAME*	*ADRESSE)*
500	**Müller**	**Köln**
550	**Meier**	**Bonn**
600	**Schulz**	**Aachen**

entlehnung

(INVNR	*LNR*	*RDAT)*
1	**550**	**300486**
2	**550**	**310586**
3	**600**	**310586**

Beispiel: Schek

<table>
<tr><th colspan="8">ABTEILUNG</th></tr>
<tr><th rowspan="3">ANR</th><th rowspan="3">ANAME</th><th colspan="2">BÚROANG</th><th colspan="4">TECH-ANG</th></tr>
<tr><th rowspan="2">BNR</th><th rowspan="2">B-ARBEIT</th><th rowspan="2">TNR</th><th rowspan="2">T-ARBEIT</th><th colspan="2">KURSE</th></tr>
<tr><th>KNR</th><th>JAHR</th></tr>
<tr><td rowspan="3">1</td><td rowspan="3">For
sch
ung</td><td rowspan="3">21
22
30
40
69</td><td rowspan="3">Bibliothek
Übersetz.
Sekretär
Patente
Sekretär</td><td>51</td><td>Programm.</td><td>1
2
5</td><td>75
76
79</td></tr>
<tr><td>52</td><td>Grundlagen
Forschung</td><td>1
2</td><td>82
79</td></tr>
<tr><td>78</td><td>Planung</td><td>2
4</td><td>76
82</td></tr>
<tr><td rowspan="3">2</td><td rowspan="3">Ent
wick
lung</td><td rowspan="3">19
25
35</td><td rowspan="3">Sekretär
Übersetz.
Hilfskraft</td><td>50</td><td>Entwurf</td><td>1
2</td><td>75
77</td></tr>
<tr><td>80</td><td>Wartung</td><td>3</td><td>82</td></tr>
<tr><td>81</td><td>Planung</td><td>2
3</td><td>81
82</td></tr>
</table>

"Zeige alle Bücher von Autor 'Date'."
E1 = $\sigma_{AUTOR\,=\,'DATE'}$ (buch)

"Zeige alle Verlage (von denen Bücher vorhanden sind)."
E2 = π_{VERLAG} (buch)

"Zeige die Rückgabedaten der Bücher mit den Inventarnummern 1 bzw. 2.
E3 = π_{RDAT} ($\sigma_{INVNR\,=\,1\,\vee\,INVNR\,=2}$ (entlehnung))

"Zeige die Adressen der Leser, die zur Zeit ein Buch entliehen haben."
E4 = $\pi_{ADRESSE}$ (person $\bowtie$ entlehnung)

"Zeige Buchtitel und Lesernamen jedes aktuellen Entleihvorgangs."
E5 = $\pi_{TITEL,NAME}$ (buch $\bowtie$ person $\bowtie$ entlehnung)

Nest-Operator:

R

A	B	C	D
a1	b1	c2	d1
a1	b1	c3	d2
a2	b1	c3	d2
a2	b1	c1	d1
a2	b2	c4	d2

n[C, D:CD] (R)

R

A	B	CD: C	CD: D
a1	b1	c2 c3	d1 d2
a2	b1	c3 c1	d2 d1
a2	b2	c4	d2

Unnest-Operator: invers

R

A	B	C	D
a1	b1	c2	d1
a1	b1	c3	d2
a2	b1	c3	d2
a2	b1	c1	d1
a2	b2	c4	d2

un[C,D:CD] (R)

R

A	B	CD: C	CD: D
a1	b1	c2 c3	d1 d2
a2	b1	c3 c1	d2 d1
a2	b2	c4	d2