COSC265 — Relational Database Systems

Neville Churcher

Department of Computer Science & Software Engineering
University of Canterbury

2021

Relational Data Model

Domain: Set (finite?) of values

Relation: Subset of the Cartesian product of one or more domains

Tuple: Elements (members) of a relation

$$r = \{t | t \in \otimes \mathcal{D}_i\}$$

are k-tuples

$$\{(\nu_1, \nu_2 \ldots \nu_k) | \nu_i \in \mathcal{D}_i\}$$

Tables: Relations can be thought of as tables whose (unordered) rows are tuples and whose columns are components

Attributes

- Names for tuple components (columns).
- Tuples viewed as mappings from attributes names to values in corresponding domains
- If columns have names then order of columns does not matter (set of mappings)
- Otherwise component order does matter. The relation is viewed as a set of lists
- Both forms have uses. Conversion between them is achieved by assigning arbitrary attribute names or by fixing attribute order

Relations & Schemes

- A relation *scheme* or *schema* is the set of attributes

$$R = \{A_1, A_2 \ldots A_n\}$$

- A relation over schema R is denoted by

$$r(R) \equiv r(A_1, A_2 \ldots A_n)$$

- *Concatenation* of attributes (or sets of attributes) used to denote *union*

$$R \equiv A_1 A_2 \ldots A_n$$

- Schema extraction via

$$R = attr(r)$$

- Restriction of tuple to specified components (by name or position)

$$s[i], s \in r(R)$$

Keys & Superkeys

- A key K is a set of attributes such that

$$\nexists t_i, t_j : t_i[K] = t_j[K]$$

- Duplicate tuples not allowed as all elements of set (relation) must be distinct
- R is a key for any $r(R)$
- If K, K' are keys and $K \supseteq K'$ then K is a *superkey*
- Relation may have more than one *candidate key*
- Usually assume ‘key’ has no proper subset that is also a key
- Keys should be time ($\equiv$ tuple value) independent

Complements

- Primarily interested in *finite* relations
- Define $dom(R)$ as set of all tuples over attributes of R and their domains
- Define relation complement of $r(R)$ as

$$\bar{r} = dom(R) - r$$

- If R has any ∞ domains then $\bar{r}$ is ∞
- If $R \equiv A_1 A_2 \ldots A_n$ and $D_i \equiv dom(A_i)$ then define the *active domain* of A_i in r as

$$adom(A_i, r) = \{d \in D_i | \exists t \in r, t(A_i) = d\}$$

- Then can define $adom(R, r)$ as set of all tuples over the attributes of R and their active domains relative to R
- Active complement (always a finite relation)

$$\tilde{r} = adom(R, r) - r$$

Example

$$R = ABC, dom(A) = \{a_1, a_2\}, dom(B) = \{b_1, b_2, b_3\}, dom(C) = \{c_1, c_2\}$$

$r(R)$

A	B	C
a_1	b_1	c_1
a_1	b_2	c_1
a_2	b_1	c_2

$dom(R)$

A	B	C
a_1	b_1	c_1
a_1	b_1	c_2
a_1	b_2	c_1
a_1	b_2	c_2
a_1	b_3	c_1
a_1	b_3	c_2
a_2	b_1	c_1
a_2	b_1	c_2
a_2	b_2	c_1
a_2	b_2	c_2
a_2	b_3	c_1
a_2	b_3	c_2

$\bar{r} = dom(R) - r$

A	B	C
a_1	b_1	c_2
a_1	b_2	c_2
a_1	b_3	c_1
a_1	b_3	c_2
a_2	b_1	c_1
a_2	b_2	c_1
a_2	b_2	c_2
a_2	b_3	c_1
a_2	b_3	c_2

$\tilde{r} = adom(R) - r$

A	B	C
a_1	b_1	c_2
a_1	b_2	c_2
a_2	b_1	c_1
a_2	b_2	c_1
a_2	b_2	c_2

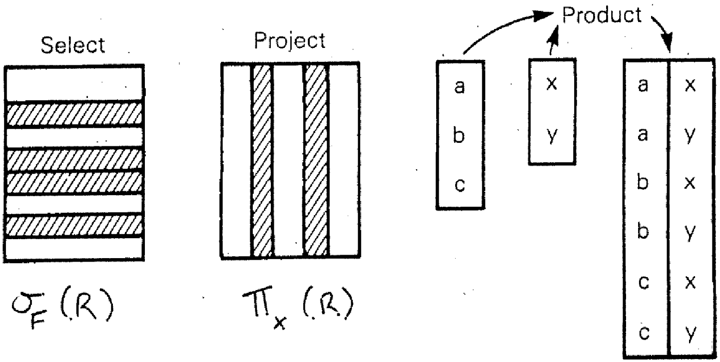
Observations

- Active complements can arise naturally as in *trained_in(Employee, Dept)*
- If every employee (and dept) appears in at least one tuple, then tuples of *trained_in* are those combinations which have *not* taken place
- Thought for the day: If $\|\tilde{r}\| < \|r\|$ then could the active complement be used as a storage compression device?

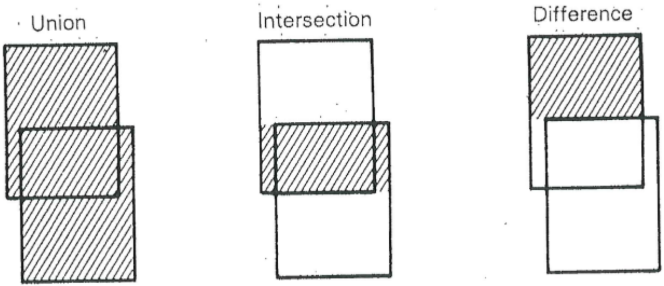
Relational Algebra

- ☆ Operands are constant relations or variables denoting relations of fixed 'arity'
- ☆ Relations on the same scheme are sets over the same universe
- ☆ Set of all relations exhibits closure under the operators of the relational algebra
  - $\cup$ Union
  - $\cap$ Intersection
  - $-$ Difference ('but not'). $\setminus$ also used
  - $\otimes$ Cartesian product
  - σ selection
  - π Projection
  - $\bowtie$ Join
  - $\div$ Division
- ☆ All operations expressible in terms of $\{\cup, -, \otimes, \sigma, \pi\}$
- ☆ Some operators ($\cup, \cap, -$) require *union-compatible* operands (same degree, corresponding domains)

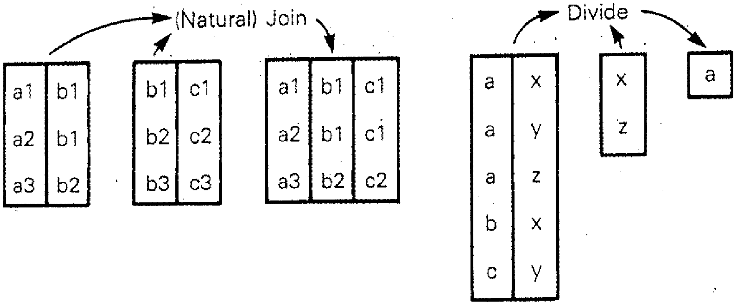
Relational Operations Illustrated



Relational Operations Illustrated



Relational Operations Illustrated



Joins

Several join operators defined as shorthand for various restrictions of the Cartesian product

Theta Joins: If r is of arity k and θ is a comparison operator, the theta join on columns i, j is written as

$$r \underset{i\theta j}{\bowtie} s \equiv \sigma_{i\theta(k+j)}(r \otimes s)$$

Equijoin This is the case where θ is '$=$'. The resulting relation has two identical (sets of) columns — so we ll usually want remove one.

Natural Join $\bowtie$ Like an equijoin, but one of the duplicate (sets of) columns removed. Only applicable when columns have names (attributes). Given $r(R)$, $s(S)$ and $I = R \cap S$

$$r \bowtie s = \pi_{RS}\sigma_{r.I=s.I}(r \otimes s)$$

Theta Join Example

Given $R = ABC$, $S = DE$

$r(R)$

A	B	C
1	2	3
4	5	6
7	8	9

$s(S)$

D	E
3	1
6	2

$r \underset{B<D}{\bowtie} s$

A	B	C	D	E
1	2	3	3	1
1	2	3	6	2
4	5	6	6	2

$r \otimes s$

A	B	C	D	E
1	2	3	3	1
1	2	3	6	2
4	5	6	3	1
4	5	6	6	2
7	8	9	3	1
7	8	9	6	2

Natural Join Example

$r(R)$

A	B	C
a	b	c
d	b	c
b	b	f
c	a	d

$s(S)$

B	C	D
b	c	d
b	c	e
a	d	b

$r \bowtie s$

A	B	C	D
a	b	c	d
a	b	c	e
d	b	c	d
d	b	c	e
c	a	d	b

$r \otimes s$

A	B	C	B'	C'	D
a	b	c	b	c	d
a	b	c	b	c	e
a	b	c	a	d	b
d	b	c	b	c	d
d	b	c	b	c	e
d	b	c	a	d	b
...	...	...	...	...	...
c	a	d	a	d	b

$r \underset{r.B=s.B \wedge r.C=s.C}{\bowtie} s$

A	B	C	B'	C'	D
a	b	c	b	c	d
a	b	c	b	c	e
d	b	c	b	c	d
d	b	c	b	c	e
...	...	...	...	...	...
c	a	d	a	d	b

Relational Division Operator

Consider relations $r(R), s(S)$ where $S \subseteq R$.

Tuples $\eta \in (r \div s)$ appear in r in combination with *every* tuple in s and the schema of $r \div s$ is $R - S$

$r \div s = \{\eta | (\forall \mu \in s)(\exists \nu \in r)\, \nu[S] = \mu[S] \wedge \nu[R - S] = \eta[R - S]\}$

The tuples satisfying the second part of the definition (i.e. possible η) are given by

$$\pi_{R-S}(r)$$

The combination of all pairs of tuples from $\pi_{R-S}(r)$ and s is given by

$$\pi_{R-S}(r) \otimes s$$

Those combinations which do not appear in r (thus violating the first part of the above definition) are given by

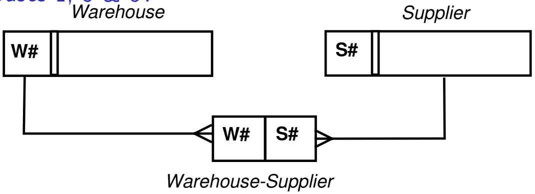
$$(\pi_{R-S}(r) \otimes s) - r$$

Thus

$$r \div s = \pi_{R-S}(r) - \pi_{R-S}((\pi_{R-S}(r) \otimes s) - r)$$

Division Example

Who supplies (each of) warehouses 1, 3 & 5?



W-S

W#	S#
W1	S1
W1	S2
W1	S3
W2	S3
W3	S1
W3	S2
W5	S1
W5	S2
W5	S4
W6	S2

$\equiv$

W-S'

W#	S#
W1	S1
W3	S1
W5	S1
W1	S2
W3	S2
W5	S2
W2	S3
W5	S4
W6	S2
W1	S3

$\div$

W135

W#
W1
W3
W5

$=$

S135

S#
S1
S2

Additional Operators

Assignment Operator($\leftarrow$)

- ☆ Not *essential* — adds no expressive power
- ☆ Convenient for complex queries involving intermediate variables/expressions

Example

$$r \leftarrow \sigma_{x<42}(s \bowtie p)$$

Renaming Operator(δ)

- ☆ Used to construct aliases
- ☆ e.g.foreign keys, distinguishing between duplicate columns in joins

Example

If $R = \ldots X \ldots$ and $r = r(R)$

$$r' = \delta_{X \leftarrow Y}(r) \equiv r'(R')$$

where $R' = \ldots Y \ldots$