Relational Algebra

Carla Teixeira Lopes

Bases de Dados
Mestrado Integrado em Engenharia Informática e Computação, FEUP

Based on Jennifer Widom slides

Division operator

Operator: /

Identifies the attribute values from a relation that are paired with **all** of the values from another relation

The division is to the Cartesian product (x) what the division is to multiplication in arithmetic

Necessary to answer queries with “all”

Cross-product and division

Division is the opposite of the cross-product

R1

A
4
8

R2

B
3
1
7

R3 = R1 x R2

R3

A	B
4	3
4	1
4	7
8	3
8	1
8	7

R3 / R2 = R1

R3 / R1 = R2

How to divide

R

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

/

S

C	D
c	d
e	f

= ?

Reorder the columns in R so the last ones are the ones in S

Order tuples in R by the first columns

Each R sub-tuple is part of the result if the sub-tuple of the last columns contains the divisor

How to divide

Reorder the columns in R so the last ones are the ones in S

R

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

/

S

C	D
c	d
e	f

= ?

How to divide

Order R by the first columns

R

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

/

S

C	D
c	d
e	f

= ?

R

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

How to divide

Each R sub-tuple is part of the result if the sub-tuple of the last columns contains the divisor

R

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

/

S

C	D
c	d
e	f

=

R/S

A	B
a	b
e	d

Example

Which members are enrolled in all sports?

EnrolledIn

Member	Sport	Payment
6078	GM	25
5819	KB	30
4526	KB	30
4526	SW	20
3955	KB	30
3955	SW	20
3955	GM	25
9876	KB	0

Sports

ID	Name
KB	Kickbox
SW	Swimming
GM	Gimnastics

Example

Which members are enrolled in all sports?

EnrolledIn

Member	Sport	Payment
6078	GM	25
5819	KB	30
4526	KB	30
4526	SW	20
3955	KB	30
3955	SW	20
3955	GM	25
9876	KB	0

Sports

ID	Name
KB	Kickbox
SW	Swimming
GM	Gimnastics

$A = \pi_{Member,Sport}\ EnrolledIn$

Member	Sport
6078	GM
5819	KB
4526	KB
4526	SW
3955	KB
3955	SW
3955	GM
9876	KB

$B = \pi_{ID}\ Sports$

ID
KB
SW
GM

A/B

Member
3955

Division doesn't add expressive power

$R(a_1, ..., a_n, b_1, ..., b_m)$

$S(b_1, ..., b_m)$

$$R/S = \Pi_{a1,...,an}(R) - \Pi_{a1,...,an}[(\Pi_{a1,...,an}(R) \times S) - R]$$

Division doesn't add expressive power

$$R/S = \Pi_{a1,\dots,an}(R) - \Pi_{a1,\dots,an}[(\Pi_{a1,\dots,an}(R) \times S) - R]$$

R

Member	Sport
5819	KB
4526	KB
4526	SW
3955	KB
3955	SW
3955	GM

S

ID
KB
SW
GM

$\Pi_{a1,\dots,an}(R)$

Member
5819
4526
3955

All tuples from the first n attributes of R

Division doesn't add expressive power

$$R/S = \Pi_{a1,\dots,an}(R) - \Pi_{a1,\dots,an}[(\Pi_{a1,\dots,an}(R) \times S) - R]$$

$\Pi_{a1,\dots,an}(R)$

Member
5819
4526
3955

S

ID
KB
SW
GM

$\Pi_{a1,\dots,an}(R) \times S$

Member	ID
5819	KB
5819	SW
5819	GM
4526	KB
4526	SW
4526	GM
3955	KB
3955	SW
3955	GM

All combinations of the first n attributes of R with the tuples of S

Division doesn't add expressive power

$$R/S = \Pi_{a1,\ldots,an}(R) - \Pi_{a1,\ldots,an}[(\Pi_{a1,\ldots,an}(R) \times S) - R]$$

$\Pi_{a1,\ldots,an}(R) \times S$

Member	ID
5819	KB
5819	SW
5819	GM
4526	KB
4526	SW
4526	GM
3955	KB
3955	SW
3955	GM

R

Member	Sport
5819	KB
4526	KB
4526	SW
3955	KB
3955	SW
3955	GM

$\Pi_{a1,\ldots,an}(R) \times S - R$

Member	ID
5819	SW
5819	GM
4526	GM

All combinations of the first n attributes of R with the tuples of S excluding the tuples that are present in R

Division doesn't add expressive power

$$R/S = \Pi_{a1,\dots,an}(R) - \Pi_{a1,\dots,an}[(\Pi_{a1,\dots,an}(R) \times S) - R]$$

$\Pi_{a1,\dots,an}(R) \text{ x } S - R$

Member	ID
5819	SW
5819	GM
4526	GM

$\Pi_{a1,\dots,an}[(\Pi_{a1,\dots,an}(R) \times S) - R]$

Member
5819
4526

First n attributes of all combinations of the first n attributes of R with the tuples of S excluding the tuples that are present in R

R

Member	Sport
5819	KB
4526	KB
4526	SW
3955	KB
3955	SW
3955	GM

S

ID
KB
SW
GM

Division doesn't add expressive power

$$R/S = \Pi_{a1,\dots,an}(R) - \Pi_{a1,\dots,an}[(\Pi_{a1,\dots,an}(R) \times S) - R]$$

R

Member	Sport
5819	KB
4526	KB
4526	SW
3955	KB
3955	SW
3955	GM

$\Pi_{a1,\dots,an}[(\Pi_{a1,\dots,an}(R) \times S) - R]$

Member
5819
4526

$\Pi_{a1,\dots,an}(R) - \Pi_{a1,\dots,an}[(\Pi_{a1,\dots,an}(R) \times S) - R]$

Member
3955

Rename operator

Changes the schema, not the instance

General Notation

$\rho_{R(A_1,\ldots,A_n)}(E)$

Abbreviated notation to only change the relation name

$\rho_R(E)$

Abbreviated notation to only change attribute names

$\rho_{A_1,\ldots,A_n}(E)$

Example

Student

sID	sName	GPA	HS

$$\rho_{Student2(ID,Name,Grade,HighSchool)}(Student)$$

Student2

ID	Name	Grade	HighSchool

Rename operator in use

To unify schemas for set operators

List of colleges and student names

$$\rho_{C(name)}(\pi_{cName} College) \cup \rho_{C(name)}(\pi_{sName} Student)$$

College

cName	state	enr

Student

sID	sName	GPA	HS

Rename operator in use

College

cName	state	enr

For disambiguation in "self-joins"

Pairs of colleges in the same state

$\sigma_{state=state}(College \times College)$?

$\sigma_{s1=s2}(\rho_{c_1(n_1,s_1,e_1)}(College) \times \rho_{c_2(n_2,s_2,e_2)}(College))$

$\rho_{c_1(n_1,s_1,e_1)}(College) \bowtie \rho_{c_2(n_2,s_2,e_2)}(College)$?

$\rho_{c_1(n_1,s,e_1)}(College) \bowtie \rho_{c_2(n_2,s,e_2)}(College)$

Rename operator in use

Pairs of **different** colleges in the same state

$$\sigma_{n_1 \neq n_2}(\rho_{c_1(n_1,s,e_1)}(College) \bowtie \rho_{c_2(n_2,s,e_2)}(College))$$

~~MIT MIT~~

Pairs of **different** colleges in the same state without repeated pairs

$$\sigma_{n_1 > n_2}(\rho_{c_1(n_1,s,e_1)}(College) \bowtie \rho_{c_2(n_2,s,e_2)}(College))$$

Stanford Berkeley
~~Berkeley Stanford~~

College

cName	**state**	**enr**
MIT	Massachusetts	NULL
Stanford	California	NULL
Berkeley	California	NULL

Agenda

~~Introduction to Relational Algebra~~

~~Operators~~

Alternate notations

Extensions to Relational Algebra

Assignment statements

Pairs of different colleges in the same state

$$\sigma_{n_1<n_2}(\rho_{c_1(n_1,s,e_1)}(College) \bowtie \rho_{c_2(n_2,s,e_2)}(College))$$

College

cName	state	enr

Alternative notation

$$C1 := \rho_{c_1(n_1,s,e_1)}(College)$$

$$C2 := \rho_{c_2(n_2,s,e_2)}(College)$$

$$CP := C1 \bowtie C2$$

$$Answ := \sigma_{n_1<n_2} CP$$

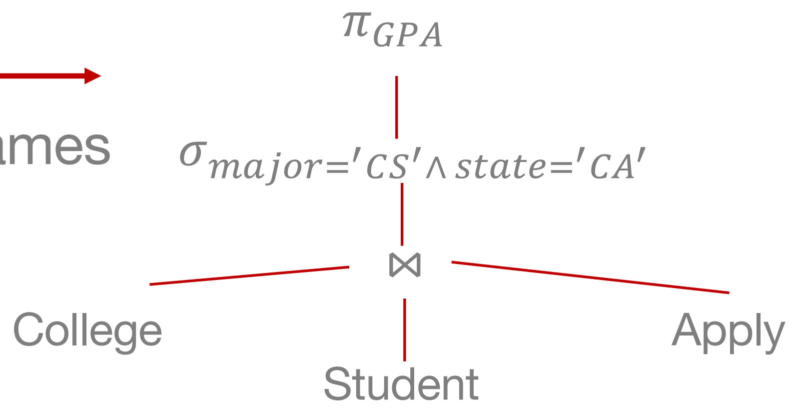
Expression trees

GPAs of students applying to CS in CA

$$\pi_{GPA}(\sigma_{major='CS' \wedge state='CA'}(Student \bowtie Apply \bowtie College))$$

Alternative notation

The leaves are always relation names



College

cName	state	enr

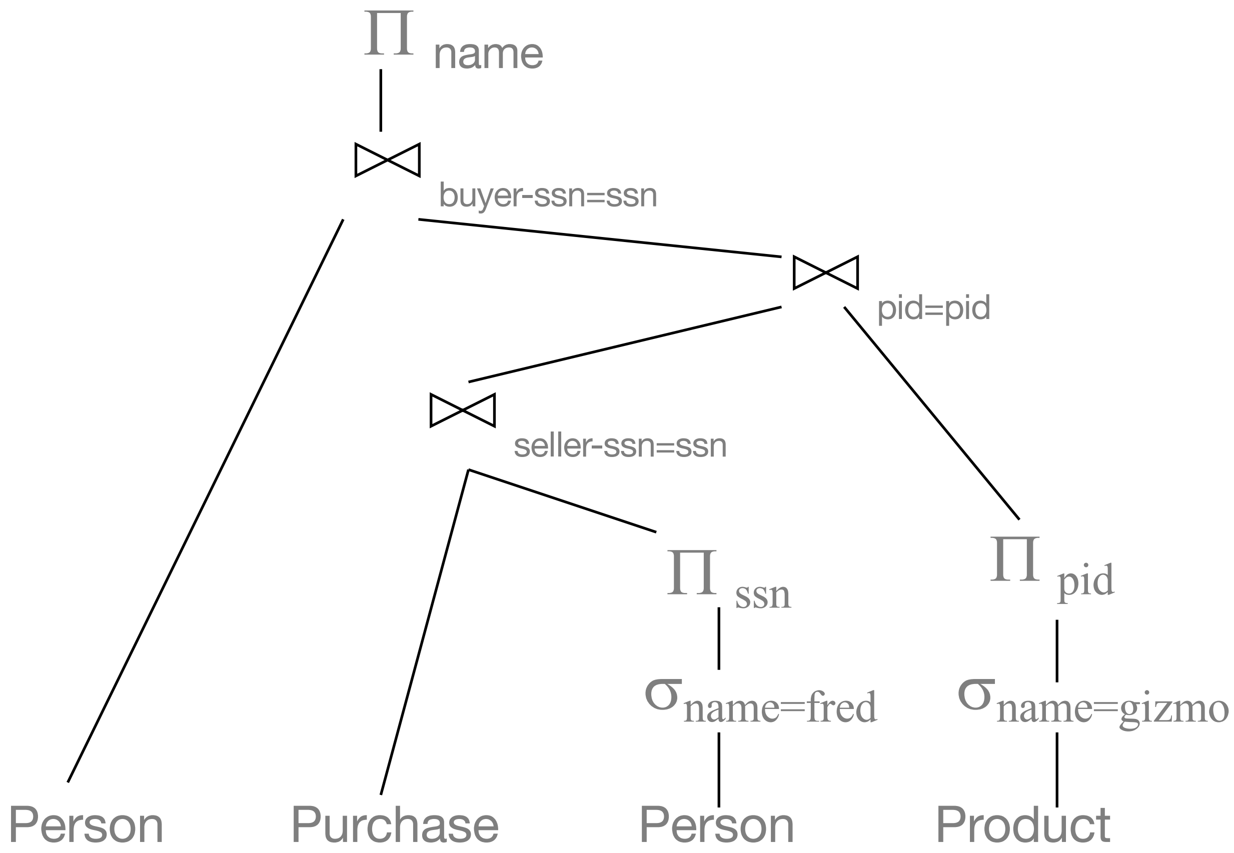
Student

sID	sName	GPA	HS

Apply

sID	cName	major	dec

Expressions can get complex



Agenda

~~Introduction to Relational Algebra~~

~~Operators~~

~~Alternate notations~~

Extensions to Relational Algebra

Extensions to Relational Algebra

Arithmetic expressions in projections

Student

sID	sName	GPA	HS
12	Mary	3.5	90
23	John	3.8	50
31	Jane	3.9	1000

List the student ids with the GPA increased in 0.2

$$\pi_{sID, new=GPA+0.2}\ Student$$

In the parameters of the projection

- Left side can only be the name of a new attribute
- Right side can only involve attributes of the involved relation

Extensions to Relational Algebra

Aggregation operators

cnt, sum, avg, max, min

$\pi_{cnt(*)}\ R$

Relation with 1 tuple and 1 attribute containing the number of tuples in R

College

cName	state	enr
MIT	Massachusetts	30000
Stanford	California	20000
Berkeley	California	10000
Harvard	Massachusetts	NULL

$\pi_{cnt(*)}\ College$

cnt(*)
4

Extensions to Relational Algebra

$\pi_{cnt(B)}\ R$

Relation with 1 tuple and 1 attribute containing the number of tuples in R with non-null values in B

College

cName	state	enr
MIT	Massachusetts	30000
Stanford	California	20000
Berkeley	California	10000
Harvard	Massachusetts	NULL

$\pi_{cnt(enr)}\ College$

cnt(enr)
3

Extensions to Relational Algebra

$\pi_{max(B)}\ R$

Relation with 1 tuple and 1 attribute containing the maximum value in B

College

cName	state	enr
MIT	Massachusetts	30000
Stanford	California	20000
Berkeley	California	10000
Harvard	Massachusetts	NULL

$\pi_{max(enr)}\ College$

max(enr)
30000

Extensions to Relational Algebra

$\pi_{A,\max(B)}\ R$

Relation with 1 tuple per each value of A and 2 attributes: the value of A and the maximum value of B for that value of A

College

cName	state	enr
MIT	Massachusetts	30000
Stanford	California	20000
Berkeley	California	10000
Harvard	Massachusetts	NULL

$\pi_{state,max(enr)}\ College$

state	max(enr)
Massachusetts	30000
California	20000

Relational Algebra Operators Summary

Core operators

R

$\sigma_{condition} E$

$\pi_{A_1,\ldots,A_n} E$

$E_1 \times E_2$

$E_1 \cup E_2$

$E_1 - E_2$

$\rho_{R(A_1,\ldots,A_n)}(E)$

Derived operators

$E_1 \bowtie E_2$

$E_1 \bowtie_\theta E_2$

$E_1 \cap E_2$

$E_1 \ltimes E_2$

$E_1 \,/\, E_2$

Parentheses are used for disambiguation

Kahoot time!

Any doubts?

Readings

Jeffrey Ullman, Jennifer Widom, A first course in Database Systems 3rd Edition

- Section 2.4 – An Algebraic Query Language
- Section 5.2 – Extended Operators of Relational Algebra