



Relational Algebra

Chapter 4

Formal Query Languages

- Foundation for commercial query languages like SQL
- Two types
 - Declarative: **Relational Calculus**
 - Describe what a user wants, rather than how to compute it
 - Procedural: **Relational Algebra**
 - Operational, very useful for representing execution plans
- Query Languages **!=** programming languages
 - QLs not expected to be “Turing complete”
 - QLs not intended to be used for complex calculations
 - QLs support easy, efficient access to large data sets

Understanding Algebra & Calculus is key to understanding SQL query processing!

Relational Algebra Preliminaries

- Query:
 - Input: Relational instances
 - Output: Relational instances
- Specified using the schemas
 - May produce different results for different instances
 - But schema of the result is fixed

Relational Operators

- Selection σ : Selects a subset of rows from relation
 - Projection π : Deletes unwanted columns from relation
 - Cross-product \times : Allows us to combine two relations
 - Set-difference $-$: Tuples in reln. 1, but not in reln. 2
 - Union \cup : Tuples in reln. 1 and in reln. 2
- Diagrammatic grouping of the first five items:
- Selection σ and Projection π are grouped by a bracket labeled "1 reln".
 - Cross-product \times , Set-difference $-$, and Union \cup are grouped by a bracket labeled "2 reln".
- Additional operations (constructed from basic ops – not essential, but very useful):
 - Intersection, **Join**, Division, Renaming
 - Because algebra is closed, we can **compose** operators

Selection



- Retrieve **rows** that satisfy a logical condition

$$\sigma_{condition}(relation)$$

- Conditions are boolean combinations (\wedge , \vee) of terms
- Each term has one of the following forms:
 - Attr1 <oper> Attr2
 - Attr <oper> valuewhere <oper> in $\{>, \geq, =, \leq, <\}$

Athlete

aid	name	sport	country
1	Mary Lou Retton	gymnastics	USA
2	Jackie Joyner-Kersey	track	USA
3	Michael Phelps	swimming	USA
4	Johann Koss	skating	Norway
5	Natalie Coughlin	swimming	USA
6	Gabby Douglas	gymnastics	USA



$$\sigma_{sport='gymnastics' \wedge country='USA'}(Athlete)$$

aid	name	sport	country
1	Mary Lou Retton	gymnastics	USA
6	Gabby Douglas	gymnastics	USA

Projection



- Selects columns
- Delete attributes that are not in projection list
- Remove **duplicates**!
 - Remember classical relational model is a **SET**

Athlete

aid	name	sport	country
1	Mary Lou Retton	gymnastics	USA
2	Jackie Joyner-Kersey	track	USA
3	Michael Phelps	swimming	USA
4	Johann Koss	skating	Norway
5	Natalie Coughlin	swimming	USA
6	Gabby Douglas	gymnastics	USA

$\pi_{sport, country}(Athlete)$

sport	country
gymnastics	USA
track	USA
swimming	USA
skating	Norway

Set Operations: Union (\cup), Intersection(\cap), Set-Difference ($-$)



- **Input:** Two **union-compatible** relations
 - Same number and type of attributes, in same order
- **Field names of result:** uses the name from the FIRST input

name
Mary Lou Retton
Jackie Joyner-Kersey
Michael Phelps
Johann Koss
Natalie Coughlin
Gabby Douglas

—

AthleteName
Mary Lou Retton
Jackie Joyner-Kersey
Apollo Ono
Picabo Street
Natalie Coughlin
Bode Miller

=

name
Michael Phelps
Johann Koss
Gabby Douglas

Cross-Product (Cartesian Product)



- **Result Schema:**

- All fields from both relations, names inherited if possible

Athlete

aid	name	sport	country
1	Mary Lou Retton	gymnastics	USA
2	Jackie Joyner-Kersee	track	USA



Venue

vid	venue
1	Los Angeles
2	Barcelona



aid	name	sport	country	vid	venue
1	Mary Lou Retton	gymnastics	USA	1	Los Angeles
2	Jackie Joyner-Kersee	track	USA	1	Los Angeles
1	Mary Lou Retton	gymnastics	USA	2	Barcelona
2	Jacki Joyner-Kersee	track	USA	2	Barcelona

Cross-Product (Cartesian Product)



- **Result Schema:**

- All fields from both relations, names inherited if possible
- If both input relations have a field with the same name, use the rename operator ρ

Athlete

aid	name	sport	venue
1	Mary Lou Retton	gymnastics	USA
2	Jackie Joyner-Kersee	track	USA



Venue

vid	venue
1	Los Angeles
2	Barcelona



aid	name	sport	(venue)	vid	(venue)
1	Mary Lou Retton	gymnastics	USA	1	Los Angeles
2	Jackie Joyner-Kersee	track	USA	1	Los Angeles
1	Mary Lou Retton	gymnastics	USA	2	Barcelona
2	Jacki Joyner-Kersee	track	USA	2	Barcelona

Rename Operator



$\rho_{AFTER}(BEFORE)$ or $\rho(AFTER, BEFORE)$

- Use cases:
 - Shorthand
 - Self-joins
 - If two relations have the same field

- Example: Athlete x Athlete

$\rho(A1, Athlete) \times \rho(A2, Athlete)$

Athlete

aid	name	sport	country
1	Mary Lou Retton	gymnastics	USA
2	Jackie Joyner-Kersey	track	USA

Derived Operators: Joins

- Most common way of combining information from two tables
- **Conditional join or Θ -join**
 - $R \bowtie_c S = \sigma_c(R \times S)$, where c is a condition
- **Equijoin**
 - Join condition consists only of equalities
- **Natural Join**
 - Like equijoin but drops the duplicate columns
 - $R \bowtie S$
- Despite equivalence, joins faster than cross-product!

Examples: Writing Queries in RA



Sailors (sid, sname, rating, age)

Reserves (sid, bid, day)

Boats (bid, bname, color)

Sailors

sid	sname	rating	age
22	Dustin	7	45.0
29	Brutus	1	33.0
31	Lubber	8	55.5
32	Andy	8	25.5
58	Rusty	10	35.0
64	Horatio	7	35.0
71	Zorba	10	16.0
74	Horatio	9	35.0
85	Art	3	25.5
95	Bob	3	63.5

Reserves

sid	bid	day
22	101	10/10/98
22	102	10/10/98
22	103	10/8/98
22	104	10/7/98
31	102	11/10/98
31	103	11/6/98
31	104	11/12/98
64	101	9/5/98
64	102	9/8/98
74	103	9/8/98

Boats

bid	bname	color
101	Interlake	blue
102	Interlake	red
103	Clipper	green
104	Marine	red

**Find names of
sailors who have
reserved boat #103**



Find names of sailors who have reserved boat #103



Solution 1: $\pi_{sname}((\sigma_{bid=103} Reserves) \bowtie Sailors)$

Solution 2: $\rho(Temp1, \sigma_{bid=103} Reserves)$
 $\rho(Temp2, Temp1 \bowtie Sailors)$
 $\pi_{sname}(Temp2)$

Solution 3: $\pi_{sname}(\sigma_{bid=103}(Reserves \bowtie Sailors))$

Find names of sailors who have reserved a red boat



$$\pi_{sname}((\sigma_{color='red'} Boats) \bowtie Reserves \bowtie Sailors)$$

Sailors

sid	sname	rating	age
22	Dustin	7	45.0
29	Brutus	1	33.0
31	Lubber	8	55.5
32	Andy	8	25.5
58	Rusty	10	35.0
64	Horatio	7	35.0
71	Zorba	10	16.0
74	Horatio	9	35.0
85	Art	3	25.5
95	Bob	3	63.5

Reserves

sid	bid	day
22	101	10/10/98
22	102	10/10/98
22	103	10/8/98
22	104	10/7/98
31	102	11/10/98
31	103	11/6/98
31	104	11/12/98
64	101	9/5/98
64	102	9/8/98
74	103	9/8/98

Boats

bid	bname	color
101	Interlake	blue
102	Interlake	red
103	Clipper	green
104	Marine	red

Find names of sailors who have reserved a red boat



$$\pi_{sname}((\sigma_{color='red'} Boats) \bowtie Reserves \bowtie Sailors)$$

An **equivalent** solution:

$$\pi_{sname} (\pi_{sid} ((\pi_{bid} \sigma_{color='red'} Boats) \bowtie Res) \bowtie Sailors)$$

The query optimizer chooses from the (equivalent) expressions and chooses one for efficiency of evaluation.

Sailors

sid	sname	rating	age
22	Dustin	7	45.0
29	Brutus	1	33.0
31	Lubber	8	55.5
32	Andy	8	25.5
58	Rusty	10	35.0
64	Horatio	7	35.0
71	Zorba	10	16.0
74	Horatio	9	35.0
85	Art	3	25.5
95	Bob	3	63.5

Reserves

sid	bid	day
22	101	10/10/98
22	102	10/10/98
22	103	10/8/98
22	104	10/7/98
31	102	11/10/98
31	103	11/6/98
31	104	11/12/98
64	101	9/5/98
64	102	9/8/98
74	103	9/8/98

Boats

bid	bname	color
101	Interlake	blue
102	Interlake	red
103	Clipper	green
104	Marine	red



$$\pi_{sname}(\text{Reserves} \bowtie \text{Sailors})$$

Sailors

sid	sname	rating	age
22	Dustin	7	45.0
29	Brutus	1	33.0
31	Lubber	8	55.5
32	Andy	8	25.5
58	Rusty	10	35.0
64	Horatio	7	35.0
71	Zorba	10	16.0
74	Horatio	9	35.0
85	Art	3	25.5
95	Bob	3	63.5

Reserves

sid	bid	day
22	101	10/10/98
22	102	10/10/98
22	103	10/8/98
22	104	10/7/98
31	102	11/10/98
31	103	11/6/98
31	104	11/12/98
64	101	9/5/98
64	102	9/8/98
74	103	9/8/98

Boats

bid	bname	color
101	Interlake	blue
102	Interlake	red
103	Clipper	green
104	Marine	red

Find names of sailors who
have reserved at least one boat



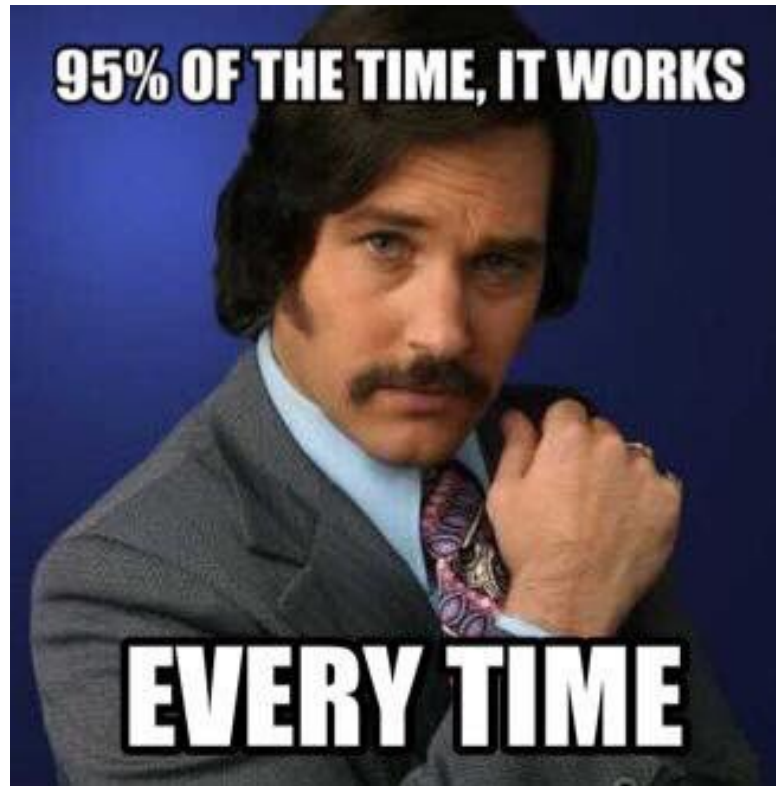
$$\pi_{sname}(\text{Reserves} \bowtie \text{Sailors})$$

Sailor appears in this intermediate relation
only if there is at least one Reserves tuple
with same sid.

Derived Operators: Division



- Useful for queries like:
“Find the sailors who have reserved **all** boats.”

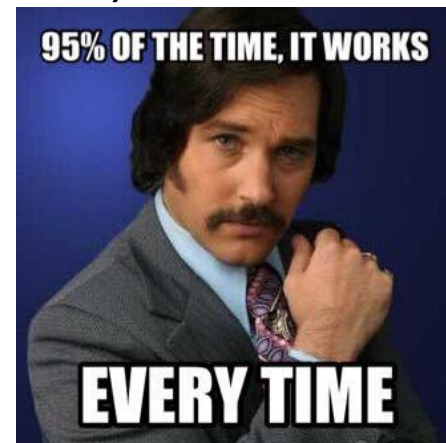


Derived Operators: Division



- Useful for queries like:

“Find the sailors who have reserved **all** boats.”
- Let A have 2 fields, x and y ; B have only field y :
 - $A/B = \{ \langle x \rangle \mid \forall \langle y \rangle \in B, \langle x, y \rangle \in A \}$
 - all x tuples (sailors) s.t. for **every** y tuple (boat) in B , there is an $\langle x, y \rangle$ tuple in A



Examples of Division A/B

$$A/B = \{ \langle x \rangle \mid \forall \langle y \rangle \in B, \langle x, y \rangle \in A \}$$

cid	bname
c1	b1
c1	b2
c1	b3
c1	b4
c2	b1
c2	b2
c3	b2
c4	b2
c4	b4

A

bname
b2

B1

cid
c1
c2
c3
c4

A/B1

bname
b2
b4

B2

cid
c1
c4

A/B2

bname
b1
b2
b4

B3

cid
c1

A/B3

Expressing A/B Using Basic Operators

- Can be equivalently expressed using basic operators
- **Idea:** For A/B, compute all x values that are not disqualified by some y value in B
 - x value is **disqualified** if by attaching y value from B, we obtain an $\langle x, y \rangle$ tuple that is not in A

Can you express this operator using basic operators?



$$A/B: \pi_x(A) - \pi_x(\underbrace{(\pi_x(A) \times B) - A}_{\text{Disqualified } x \text{ values}})$$

Examples of Division A/B

$$\pi_x(A) - \pi_x((\pi_x(A) \times B) - A)$$

cid	bname
c1	b1
c1	b2
c1	b3
c1	b4
c2	b1
c2	b2
c3	b2
c4	b2
c4	b4

A

bname
b2
b4

B

cid	bname
c1	b2
c1	b4
c2	b2
c2	b4
c3	b2
c3	b4
c4	b2
c4	b4

-A

cid
c1
c4

A/B

$$\pi_x(A) \times B$$

Find names of sailors who have reserved a red OR green boat



- Identify all red or green boats, then
- find sailors who have reserved one of these boats

Sailors

sid	sname	rating	age
22	Dustin	7	45.0
29	Brutus	1	33.0
31	Lubber	8	55.5
32	Andy	8	25.5
58	Rusty	10	35.0
64	Horatio	7	35.0
71	Zorba	10	16.0
74	Horatio	9	35.0
85	Art	3	25.5
95	Bob	3	63.5

Reserves

sid	bid	day
22	101	10/10/98
22	102	10/10/98
22	103	10/8/98
22	104	10/7/98
31	102	11/10/98
31	103	11/6/98
31	104	11/12/98
64	101	9/5/98
64	102	9/8/98
74	103	9/8/98

Boats

bid	bname	color
101	Interlake	blue
102	Interlake	red
103	Clipper	green
104	Marine	red

Find names of sailors who have reserved a red OR green boat



- Identify all red or green boats, then
- find sailors who have reserved one of these boats:

$$\rho \text{ (Tempboats, } (\sigma_{color='red' \vee color='green'} \text{Boats}))$$

$$\pi_{sname}(\text{Tempboats} \bowtie \text{Reserves} \bowtie \text{Sailors})$$

Equivalent:

$$\rho(\text{Tempboats}, (\sigma_{color='red'}(\text{Boats}) \cup \sigma_{color='green'}(\text{Boats})))$$

$$\pi_{sname}(\text{Tempboats} \bowtie \text{Reserves} \bowtie \text{Sailors})$$

Find names of sailors who have reserved a red AND green boat



?

ρ (*Tempboats*, ($\sigma_{color='red' \text{ AND } color='green'}$ *Boats*))

$\pi_{sname}(\textit{Tempboats} \bowtie \textit{Reserves} \bowtie \textit{Sailors})$



Find names of sailors who have reserved a red AND green boat



- Identify
 - sailors who have reserved red boats
 - sailors who have reserved green boats
- Then find the intersection (sid is a key for Sailors):

Sailors

sid	sname	rating	age
22	Dustin	7	45.0
29	Brutus	1	33.0
31	Lubber	8	55.5
32	Andy	8	25.5
58	Rusty	10	35.0
64	Horatio	7	35.0
71	Zorba	10	16.0
74	Horatio	9	35.0
85	Art	3	25.5
95	Bob	3	63.5

Reserves

sid	bid	day
22	101	10/10/98
22	102	10/10/98
22	103	10/8/98
22	104	10/7/98
31	102	11/10/98
31	103	11/6/98
31	104	11/12/98
64	101	9/5/98
64	102	9/8/98
74	103	9/8/98

Boats

bid	bname	color
101	Interlake	blue
102	Interlake	red
103	Clipper	green
104	Marine	red

Find names of sailors who have reserved a red AND green boat



- Identify
 - sailors who have reserved red boats
 - sailors who have reserved green boats
- Then find the intersection (sid is a key for Sailors):

$$\rho (Tempred, \pi_{sid}((\sigma_{color='red'}Boats) \bowtie Reserves))$$

$$\rho (Tempgreen, \pi_{sid}((\sigma_{color='green'}Boats) \bowtie Reserves))$$

$$\pi_{sname}((Tempred \cap Tempgreen) \bowtie Sailors)$$

Find the SIDs of sailors over age 20 who have not reserved a red boat



$$\pi_{sid}(\sigma_{age>20}Sailors) -$$

$$\pi_{sid}((\sigma_{color='red'}Boats) \bowtie Reserves)$$

Sailors

sid	sname	rating	age
22	Dustin	7	45.0
29	Brutus	1	33.0
31	Lubber	8	55.5
32	Andy	8	25.5
58	Rusty	10	35.0
64	Horatio	7	35.0
71	Zorba	10	16.0
74	Horatio	9	35.0
85	Art	3	25.5
95	Bob	3	63.5

Reserves

sid	bid	day
22	101	10/10/98
22	102	10/10/98
22	103	10/8/98
22	104	10/7/98
31	102	11/10/98
31	103	11/6/98
31	104	11/12/98
64	101	9/5/98
64	102	9/8/98
74	103	9/8/98

Boats

bid	bname	color
101	Interlake	blue
102	Interlake	red
103	Clipper	green
104	Marine	red

Find the names of sailors who have reserved **all** boats



- Uses division; schemas of the input relations must be carefully chosen:

$$\rho (Tempsids, (\pi_{sid,bid} Reserves) / (\pi_{bid} Boats))$$

$$\pi_{sname} (Tempsids \bowtie Sailors)$$

Sailors

sid	sname	rating	age
22	Dustin	7	45.0
29	Brutus	1	33.0
31	Lubber	8	55.5
32	Andy	8	25.5
58	Rusty	10	35.0
64	Horatio	7	35.0
71	Zorba	10	16.0
74	Horatio	9	35.0
85	Art	3	25.5
95	Bob	3	63.5

Reserves

sid	bid	day
22	101	10/10/98
22	102	10/10/98
22	103	10/8/98
22	104	10/7/98
31	102	11/10/98
31	103	11/6/98
31	104	11/12/98
64	101	9/5/98
64	102	9/8/98
74	103	9/8/98

Boats

bid	bname	color
101	Interlake	blue
102	Interlake	red
103	Clipper	green
104	Marine	red

Review

- Selection
- Projection
- Cross-product
- Set-difference
- Union
- Join

Exercises 4.1, 4.3, 4.5