Relational Algebra

Trishla Shah

trishla.shah@smu.ca

Basic Concepts

Five basic operators on tables:

1. *Select* - σ_θ
2. *Project* - Π_L
3. *Union* - $\cup$
4. *Difference* – $\setminus$
5. *cross product* - $\times$

Cont…

Input: one table (σ_θ, Π_L), or
two tables ($\cup$, $\setminus$, $\times$)

Output: a table

The operators $\cup$, $\setminus$, $\times$ are the usual set operators

$T_1 \cup T_2 = \{t : t \in T_1 \text{ or } t \in T_2\}$

// Recall a table is a set of tuples, so no repeated tuples allowed.

$T_1 \setminus T_2 = \{t : t \in T_1 \text{ and } t \notin T_2\}$

$T_1 \times T_2 = \{\langle a_1,\dots,a_k,b_1,\dots,b_l\rangle : \langle a_1,\dots,a_k\rangle \in T_1 \text{ and } \langle b_1,\dots,b_l\rangle \in T_2\}$

For $T_1 \cup T_2$ and $T_1 \setminus T_2$ the tables T_1, T_2 must be compatible: same number of attributes, and corresponding attributes must have same domains.

Exercise

Suppose T_1 has n tuples and T_2 has m tuples.

- How many tuples in $T_1 \times T_2$?
- How many tuples in $T_1 \cup T_2$?
- How many tuples in $T_1 \setminus T_2$?

Thus, the cross product can result into large tables.

Answer

- Suppose T_1 has n tuples and T_2 has m tuples.

- How many tuples in $T_1 \times T_2$? Answer = n*m
- How many tuples in $T_1 \cup T_2$? Answer = between max(n,m) and n+m
- How many tuples in $T_1 \setminus T_2$? Answer: between 0 and n

Thus, the cross product can result into large tables.

Project - Π

- Projection list L = list of attributes

- $\Pi_L(T)$ = the table consisting of the columns of T whose names are listed in L

Project – Π (Example 1)

TABLE S

sno	sname	status	city
S1	Smith	20	London
S2	Jones	10	Paris
S3	Blake	30	Paris
S4	Clarke	20	London
S5	Adams	30	Athens

TABLE P

pno	pname	color	weight	city
P1	nut	red	12	London
P2	bolt	green	17	Paris
P3	screw	blue	17	Rome
P4	screw	red	14	London
P5	cam	blue	12	Paris
P6	cog	red	19	London

TABLE SP (shipments)

sno	pno	qty
S1	P1	300
S1	P2	200
S1	P3	400
S1	P4	200
S1	P5	100
S1	P6	100
S2	P1	300
S2	P2	400
S3	P2	200
S4	P2	200
S4	P4	300
S4	P5	400

$\Pi_{sno,status}(S)$ = ?

Project – Π (Example 1 - Answer)

TABLE S

sno	sname	status	city
S1	Smith	20	London
S2	Jones	10	Paris
S3	Blake	30	Paris
S4	Clarke	20	London
S5	Adams	30	Athens

TABLE P

pno	pname	color	weight	city
P1	nut	red	12	London
P2	bolt	green	17	Paris
P3	screw	blue	17	Rome
P4	screw	red	14	London
P5	cam	blue	12	Paris
P6	cog	red	19	London

TABLE SP (shipments)

sno	pno	qty
S1	P1	300
S1	P2	200
S1	P3	400
S1	P4	200
S1	P5	100
S1	P6	100
S2	P1	300
S2	P2	400
S3	P2	200
S4	P2	200
S4	P4	300
S4	P5	400

Πsno,status(S) =

S1	20
S2	10
S3	30
S4	20
S5	30

Project – Π (Example 2)

TABLE S

sno	**sname**	**status**	**city**
S1	Smith	20	London
S2	Jones	10	Paris
S3	Blake	30	Paris
S4	Clarke	20	London
S5	Adams	30	Athens

TABLE P

pno	**pname**	**color**	**weight**	**city**
P1	nut	red	12	London
P2	bolt	green	17	Paris
P3	screw	blue	17	Rome
P4	screw	red	14	London
P5	cam	blue	12	Paris
P6	cog	red	19	London

TABLE SP (shipments)

sno	**pno**	**qty**
S1	P1	300
S1	P2	200
S1	P3	400
S1	P4	200
S1	P5	100
S1	P6	100
S2	P1	300
S2	P2	400
S3	P2	200
S4	P2	200
S4	P4	300
S4	P5	400

$\Pi_{sno,status}(S)$ =

S1	20
S2	10
S3	30
S4	20
S5	30

$\Pi_{status}(S)$ = ?

Project – Π (Example 2 - Answer)

TABLE S

sno	sname	status	city
S1	Smith	20	London
S2	Jones	10	Paris
S3	Blake	30	Paris
S4	Clarke	20	London
S5	Adams	30	Athens

TABLE P

pno	pname	color	weight	city
P1	nut	red	12	London
P2	bolt	green	17	Paris
P3	screw	blue	17	Rome
P4	screw	red	14	London
P5	cam	blue	12	Paris
P6	cog	red	19	London

TABLE SP (shipments)

sno	pno	qty
S1	P1	300
S1	P2	200
S1	P3	400
S1	P4	200
S1	P5	100
S1	P6	100
S2	P1	300
S2	P2	400
S3	P2	200
S4	P2	200
S4	P4	300
S4	P5	400

$\Pi_{\text{sno,status}}(S)$ =

S1	20
S2	10
S3	30
S4	20
S5	30

$\Pi_{\text{status}}(S)$ =

20
10
30

select - σ_θ

Selection condition θ. Can:

- *Simple:* X # Y or X # c, where X and Y are attributes, c constant value, and # in {=, <, <=, >, >=, !=}
- *Composite*: (**NOT** θ_1), (θ_1 **AND** θ_2), (θ_1 **OR** θ_2), where θ_1 and θ_2 are selection conditions

$\sigma_\theta(T)$ = all tuples of T satisfying condition θ

select - σ_θ (Example 1)

TABLE S

sno	sname	status	city
S1	Smith	20	London
S2	Jones	10	Paris
S3	Blake	30	Paris
S4	Clarke	20	London
S5	Adams	30	Athens

TABLE SP (shipments)

sno	pno	qty
S1	P1	300
S1	P2	200
S1	P3	400
S1	P4	200
S1	P5	100
S1	P6	100
S2	P1	300
S2	P2	400
S3	P2	200
S4	P2	200
S4	P4	300
S4	P5	400

TABLE P

pno	pname	color	weight	city
P1	nut	red	12	London
P2	bolt	green	17	Paris
P3	screw	blue	17	Rome
P4	screw	red	14	London
P5	cam	blue	12	Paris
P6	cog	red	19	London

$\sigma_{status>20}(S)$ = ?

select - σ_θ (Example 1 - Answer)

TABLE S

sno	sname	status	city
S1	Smith	20	London
S2	Jones	10	Paris
S3	Blake	30	Paris
S4	Clarke	20	London
S5	Adams	30	Athens

TABLE P

pno	pname	color	weight	city
P1	nut	red	12	London
P2	bolt	green	17	Paris
P3	screw	blue	17	Rome
P4	screw	red	14	London
P5	cam	blue	12	Paris
P6	cog	red	19	London

TABLE SP (shipments)

sno	pno	qty
S1	P1	300
S1	P2	200
S1	P3	400
S1	P4	200
S1	P5	100
S1	P6	100
S2	P1	300
S2	P2	400
S3	P2	200
S4	P2	200
S4	P4	300
S4	P5	400

$\sigma_{status>20}(S)$ =

S3	Blake	30	Paris
S5	Adams	30	Athens

select - σ_θ (Example 2)

TABLE S

sno	sname	status	city
S1	Smith	20	London
S2	Jones	10	Paris
S3	Blake	30	Paris
S4	Clarke	20	London
S5	Adams	30	Athens

TABLE P

pno	pname	color	weight	city
P1	nut	red	12	London
P2	bolt	green	17	Paris
P3	screw	blue	17	Rome
P4	screw	red	14	London
P5	cam	blue	12	Paris
P6	cog	red	19	London

TABLE SP (shipments)

sno	pno	qty
S1	P1	300
S1	P2	200
S1	P3	400
S1	P4	200
S1	P5	100
S1	P6	100
S2	P1	300
S2	P2	400
S3	P2	200
S4	P2	200
S4	P4	300
S4	P5	400

$\sigma_{status>20}(S)$ =

S3	Blake	30	Paris
S5	Adams	30	Athens

$\Pi_{sno}(\sigma_{status>20}(S))$ = ?

select - σ_θ (Example 2)

TABLE S

sno	sname	status	city
S1	Smith	20	London
S2	Jones	10	Paris
S3	Blake	30	Paris
S4	Clarke	20	London
S5	Adams	30	Athens

TABLE P

pno	pname	color	weight	city
P1	nut	red	12	London
P2	bolt	green	17	Paris
P3	screw	blue	17	Rome
P4	screw	red	14	London
P5	cam	blue	12	Paris
P6	cog	red	19	London

TABLE SP (shipments)

sno	pno	qty
S1	P1	300
S1	P2	200
S1	P3	400
S1	P4	200
S1	P5	100
S1	P6	100
S2	P1	300
S2	P2	400
S3	P2	200
S4	P2	200
S4	P4	300
S4	P5	400

$\sigma_{status>20}(S)$ =

S3	Blake	30	Paris
S5	Adams	30	Athens

$\Pi_{sno}(\sigma_{status>20}(S))$ = S3
S5

cross product - × (Example 1)

TABLE S

sno	sname	status	city
S1	Smith	20	London
S2	Jones	10	Paris
S3	Blake	30	Paris
S4	Clarke	20	London
S5	Adams	30	Athens

TABLE P

pno	pname	color	weight	city
P1	nut	red	12	London
P2	bolt	green	17	Paris
P3	screw	blue	17	Rome
P4	screw	red	14	London
P5	cam	blue	12	Paris
P6	cog	red	19	London

TABLE SP (shipments)

sno	pno	qty
S1	P1	300
S1	P2	200
S1	P3	400
S1	P4	200
S1	P5	100
S1	P6	100
S2	P1	300
S2	P2	400
S3	P2	200
S4	P2	200
S4	P4	300
S4	P5	400

S×P = ?

cross product - × (Example 1)

TABLE S

sno	sname	status	city
S1	Smith	20	London
S2	Jones	10	Paris
S3	Blake	30	Paris
S4	Clarke	20	London
S5	Adams	30	Athens

TABLE P

pno	pname	color	weight	city
P1	nut	red	12	London
P2	bolt	green	17	Paris
P3	screw	blue	17	Rome
P4	screw	red	14	London
P5	cam	blue	12	Paris
P6	cog	red	19	London

TABLE SP (shipments)

sno	pno	qty
S1	P1	300
S1	P2	200
S1	P3	400
S1	P4	200
S1	P5	100
S1	P6	100
S2	P1	300
S2	P2	400
S3	P2	200
S4	P2	200
S4	P4	300
S4	P5	400

S×P = (30 tuples)

S1	Smith	20	London	P1	nut	red	12	London
S1	Smith	20	London	P2	bolt	green	17	Paris
...	... ...	...	... ...	... ...	... ...	... ...	...	...
S1	Smith	20	London	P6	cog	red	19	London
S2	Jones	10	Paris	P1	nut	red	12	London
S2	Jones	10	Paris	P2	bolt	green	17	Paris
...	... ...	...	... ...	... ...	... ...	... ...	...	...
...	... ...	...	... ...	... ...	... ...	... ...	...	...
S5	Adams	30	Athens	P6	cog	red	19	London

Derived Operators

- **θ-join**: $T_1 \bowtie_\theta T_2 = \sigma_\theta(T_1 \times T_2)$ where each comparison $X_1 \# X_2$ in θ is between an attribute X_1 of T_1 and X_2 of T_2.

- **Equijoin**: θ-join using only '=' comparisons

- **Natural join**: $T_1 * T_2$ = the equijoin of the two tables such that
  θ = (X=X **AND** Y=Y **AND**…)
  where X, Y,.. are the common attributes of the two tables. Moreover, only **one copy** of each common column is included in the output table.

Note: In all comparisons (X # Y), if either of X and Y is **NULL**, then (X # Y) evaluates to **false**.

θ-join- $\bowtie_\theta$ (Example 1)

TABLE S

sno	sname	status	city
S1	Smith	20	London
S2	Jones	10	Paris
S3	Blake	30	Paris
S4	Clarke	20	London
S5	Adams	30	Athens

TABLE P

pno	pname	color	weight	city
P1	nut	red	12	London
P2	bolt	green	17	Paris
P3	screw	blue	17	Rome
P4	screw	red	14	London
P5	cam	blue	12	Paris
P6	cog	red	19	London

TABLE SP (shipments)

sno	pno	qty
S1	P1	300
S1	P2	200
S1	P3	400
S1	P4	200
S1	P5	100
S1	P6	100
S2	P1	300
S2	P2	400
S3	P2	200
S4	P2	200
S4	P4	300
S4	P5	400

$S \bowtie_{city=city} P$ = ?

θ-join- $\bowtie_\theta$ (Example 1)

TABLE S

sno	sname	status	city
S1	Smith	20	London
S2	Jones	10	Paris
S3	Blake	30	Paris
S4	Clarke	20	London
S5	Adams	30	Athens

TABLE SP (shipments)

sno	pno	qty
S1	P1	300
S1	P2	200
S1	P3	400
S1	P4	200
S1	P5	100
S1	P6	100
S2	P1	300
S2	P2	400
S3	P2	200
S4	P2	200
S4	P4	300
S4	P5	400

TABLE P

pno	pname	color	weight	city
P1	nut	red	12	London
P2	bolt	green	17	Paris
P3	screw	blue	17	Rome
P4	screw	red	14	London
P5	cam	blue	12	Paris
P6	cog	red	19	London

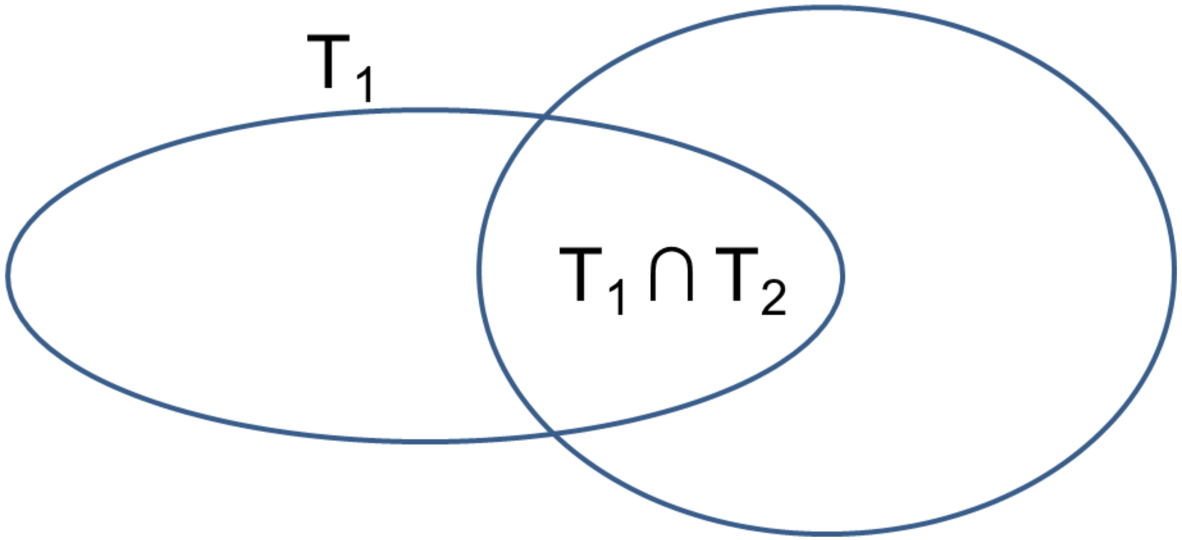
$S \bowtie_{city=city} P$ =

S1	Smith	20	London	P1	nut	red	12	London
S1	Smith	20	London	P4	screw	red	14	London
S1	Smith	20	London	P6	cog	red	19	London
S2	Jones	10	Paris	P2	bolt	green	17	Paris
S2	Jones	10	Paris	P5	cam	blue	12	Paris
S3	Blake	30	Paris	P2	bolt	green	17	Paris

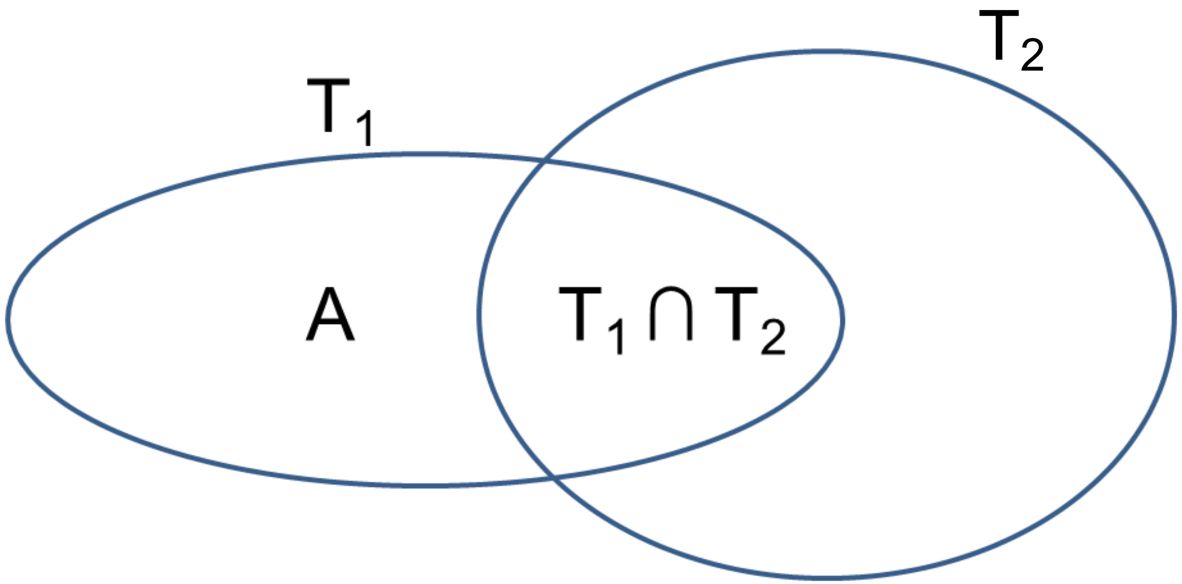
....

- Set intersection: $T_1 \cap T_2 = ?$

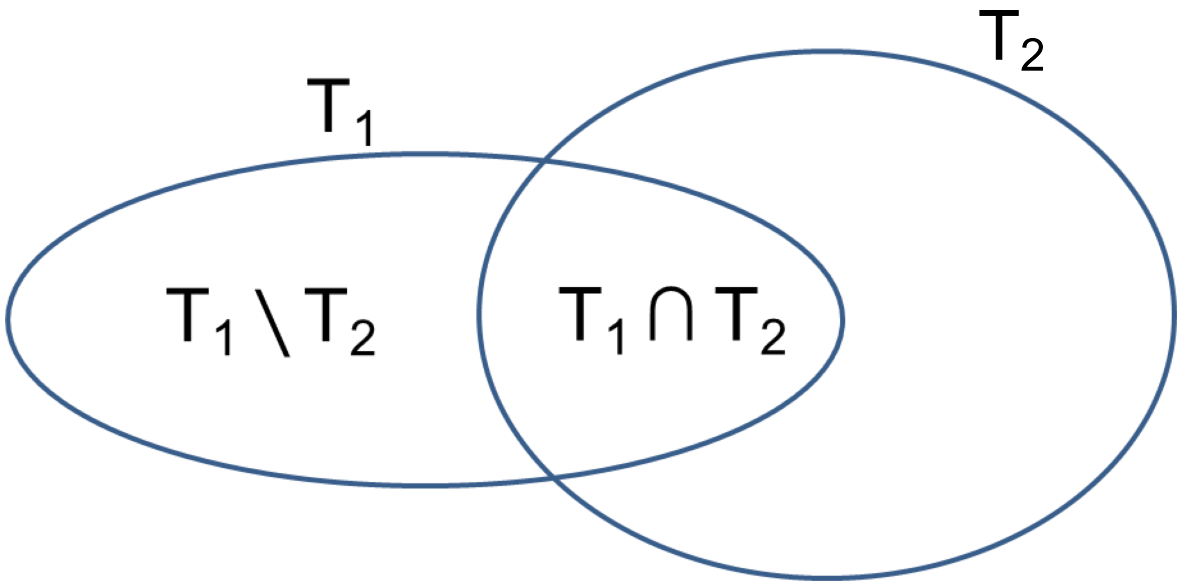
- Set intersection: $T_1 \cap T_2 = ?$



- Set intersection: $T_1 \cap T_2 = T_1 \setminus A =$?



- Set intersection: $T_1 \cap T_2 = T_1 \setminus (T_1 \setminus T_2) =$?



Exercise

TABLE S

sno	sname	status	city
S1	Smith	20	London
S2	Jones	10	Paris
S3	Blake	30	Paris
S4	Clarke	20	London
S5	Adams	30	Athens

TABLE P

pno	pname	color	weight	city
P1	nut	red	12	London
P2	bolt	green	17	Paris
P3	screw	blue	17	Rome
P4	screw	red	14	London
P5	cam	blue	12	Paris
P6	cog	red	19	London

TABLE SP (shipments)

sno	pno	qty
S1	P1	300
S1	P2	200
S1	P3	400
S1	P4	200
S1	P5	100
S1	P6	100
S2	P1	300
S2	P2	400
S3	P2	200
S4	P2	200
S4	P4	300
S4	P5	400

Queries:

- Get all suppliers located in London
- Get all locations of suppliers
- Get names of suppliers who supplied P2
- Get names of suppliers who supplied some green parts
- Get names of suppliers who didn’t supply any green parts

Exercise - Answers

TABLE S

sno	sname	status	city
S1	Smith	20	London
S2	Jones	10	Paris
S3	Blake	30	Paris
S4	Clarke	20	London
S5	Adams	30	Athens

TABLE P

pno	pname	color	weight	city
P1	nut	red	12	London
P2	bolt	green	17	Paris
P3	screw	blue	17	Rome
P4	screw	red	14	London
P5	cam	blue	12	Paris
P6	cog	red	19	London

TABLE SP (shipments)

sno	pno	qty
S1	P1	300
S1	P2	200
S1	P3	400
S1	P4	200
S1	P5	100
S1	P6	100
S2	P1	300
S2	P2	400
S3	P2	200
S4	P2	200
S4	P4	300
S4	P5	400

Queries:

- Get all suppliers located in London $Q1 = \sigma_{city="London"}(S)$
- Get all locations of suppliers $Q2 = \Pi_{city}(S)$
- Get names of suppliers who supplied P2 $Q3 = \Pi_{sname}(\sigma_{pno="P2"}(S*SP))$
- Get names of suppliers who supplied some green parts $Q4 = \Pi_{sname}(\sigma_{color="green"}(S*SP*P))$
- Get names of suppliers who didn't supply any green parts $Q5 = \Pi_{sname}(S) \setminus Q4$

Division operation (derived)

- **T1(X,Y) ÷ T2(Y)** = all tuples x such that (x, y) is in T1, for every tuple y in T2

- The two tables must have the structure shown above, otherwise division operation does not work:

- *The attributes of T2 must be a subset of those of T1*.

Cont…

T1(X, Y)	÷	T2(Y)	=	R(X)
x2 y1		y1		x1
x1 y1		y2		
x2 y2		y3		
x1 y2				
x1 y3				
x3 y1				

The value x1 is the only X-value in that is paired in T1 with every Y-value in T2.

TABLE S

sno	sname	status	city
S1	Smith	20	London
S2	Jones	10	Paris
S3	Blake	30	Paris
S4	Clarke	20	London
S5	Adams	30	Athens

TABLE P

pno	pname	color	weight	city
P1	nut	red	12	London
P2	bolt	green	17	Paris
P3	screw	blue	17	Rome
P4	screw	red	14	London
P5	cam	blue	12	Paris
P6	cog	red	19	London

TABLE SP (shipments)

sno	pno	qty
S1	P1	300
S1	P2	200
S1	P3	400
S1	P4	200
S1	P5	100
S1	P6	100
S2	P1	300
S2	P2	400
S3	P2	200
S4	P2	200
S4	P4	300
S4	P5	400

Get sno's of suppliers who have supplied **all** parts = ?

TABLE S

sno	sname	status	city
S1	Smith	20	London
S2	Jones	10	Paris
S3	Blake	30	Paris
S4	Clarke	20	London
S5	Adams	30	Athens

TABLE P

pno	pname	color	weight	city
P1	nut	red	12	London
P2	bolt	green	17	Paris
P3	screw	blue	17	Rome
P4	screw	red	14	London
P5	cam	blue	12	Paris
P6	cog	red	19	London

TABLE SP (shipments)

sno	pno	qty
S1	P1	300
S1	P2	200
S1	P3	400
S1	P4	200
S1	P5	100
S1	P6	100
S2	P1	300
S2	P2	400
S3	P2	200
S4	P2	200
S4	P4	300
S4	P5	400

Get sno's of suppliers who have supplied **all** parts = $\Pi_{sno,pno}(SP) \div \Pi_{pno}(P)$

TABLE S

sno	sname	status	city
S1	Smith	20	London
S2	Jones	10	Paris
S3	Blake	30	Paris
S4	Clarke	20	London
S5	Adams	30	Athens

TABLE P

pno	pname	color	weight	city
P1	nut	red	12	London
P2	bolt	green	17	Paris
P3	screw	blue	17	Rome
P4	screw	red	14	London
P5	cam	blue	12	Paris
P6	cog	red	19	London

TABLE SP (shipments)

sno	pno	qty
S1	P1	300
S1	P2	200
S1	P3	400
S1	P4	200
S1	P5	100
S1	P6	100
S2	P1	300
S2	P2	400
S3	P2	200
S4	P2	200
S4	P4	300
S4	P5	400

Sequences of Operations and RENAME-ing

RedOnly ← $\sigma_{color='red'}$ (**P**) // naming

Result ← Π_{sno}(SP * RedOnly) // using the new name

Sequences of Operations and RENAME-ing

- ρT(R) : renaming table R as T (same attribute names)

- *Next, assume R has n attributes*

- ρ(B1,…,Bn)(R) : renaming attributes of R as (B1,…,Bn)
- ρ T(B1,…,Bn)(R) : renaming both the table R and its attributes

TABLE S

sno	sname	status	city
S1	Smith	20	London
S2	Jones	10	Paris
S3	Blake	30	Paris
S4	Clarke	20	London
S5	Adams	30	Athens

TABLE P

pno	pname	color	weight	city
P1	nut	red	12	London
P2	bolt	green	17	Paris
P3	screw	blue	17	Rome
P4	screw	red	14	London
P5	cam	blue	12	Paris
P6	cog	red	19	London

TABLE SP (shipments)

sno	pno	qty
S1	P1	300
S1	P2	200
S1	P3	400
S1	P4	200
S1	P5	100
S1	P6	100
S2	P1	300
S2	P2	400
S3	P2	200
S4	P2	200
S4	P4	300
S4	P5	400

Sequences of Operations and RENAME-ing

Snew ← $\rho_{(sno,sname,status,\mathbf{scity})}(S)$ // S.city is called **scity** in Snew

Join1 ← Snew * SP // natural join on sno

Join2 ← Join1 * P // natural join only on pno, **city** is not common

Result ← $\Pi_{sname,pname,qty}(Join2)$

Exercise

- Show that $\div$ is indeed a derived operation.
- We compute $T1(X, Y) \div T2(Y)$ in three steps

Q1 <-- $\Pi_X(T1)$ // the X-values in T1

Q2 <-- $\Pi_X(Q1 \times T2 \setminus T1)$ // the X-values in T1 not paired

// …with every Y-value in T2

Q3 <-- $Q1 \setminus Q2$ // Q3 is $T1(X, Y) \div T2(Y)$

Quick Note About SQL: The 5 basic operations

$\Pi_{A,B}(R)$: SELECT A, B FROM R ;

$\sigma_\theta(R)$: SELECT * FROM R WHERE Θ ;

R U S: (SELECT * FROM R) UNION
(SELECT * FROM S) ;

R – S: (SELECT * FROM R) EXCEPT
(SELECT * FROM S) ;

R x S: SELECT * FROM R, S ;