Querying Relations - **Division**

pm jat @ daiict

Supply Parts Database

- Suppliers(sid, sname, city)
- Parts(pid, pname, color)
- Supplies(sid, pid, cost)

suppliers

sid [PK] integer	sname character var	city character varying
101	ABC	Mumbai
102	PQR	Delhi
103	XYZ	Ahmedabad

supplies

sid	pid	cost
101	1	100
102	1	120
101	3	160
103	2	210
102	2	220
102	3	150
102	4	400
102	5	500

parts

pid	pname	color
1	PART-1	RED
2	PART-2	GREEN
3	PART-3	RED
4	PART-4	BLUE
5	PART-5	GREEN

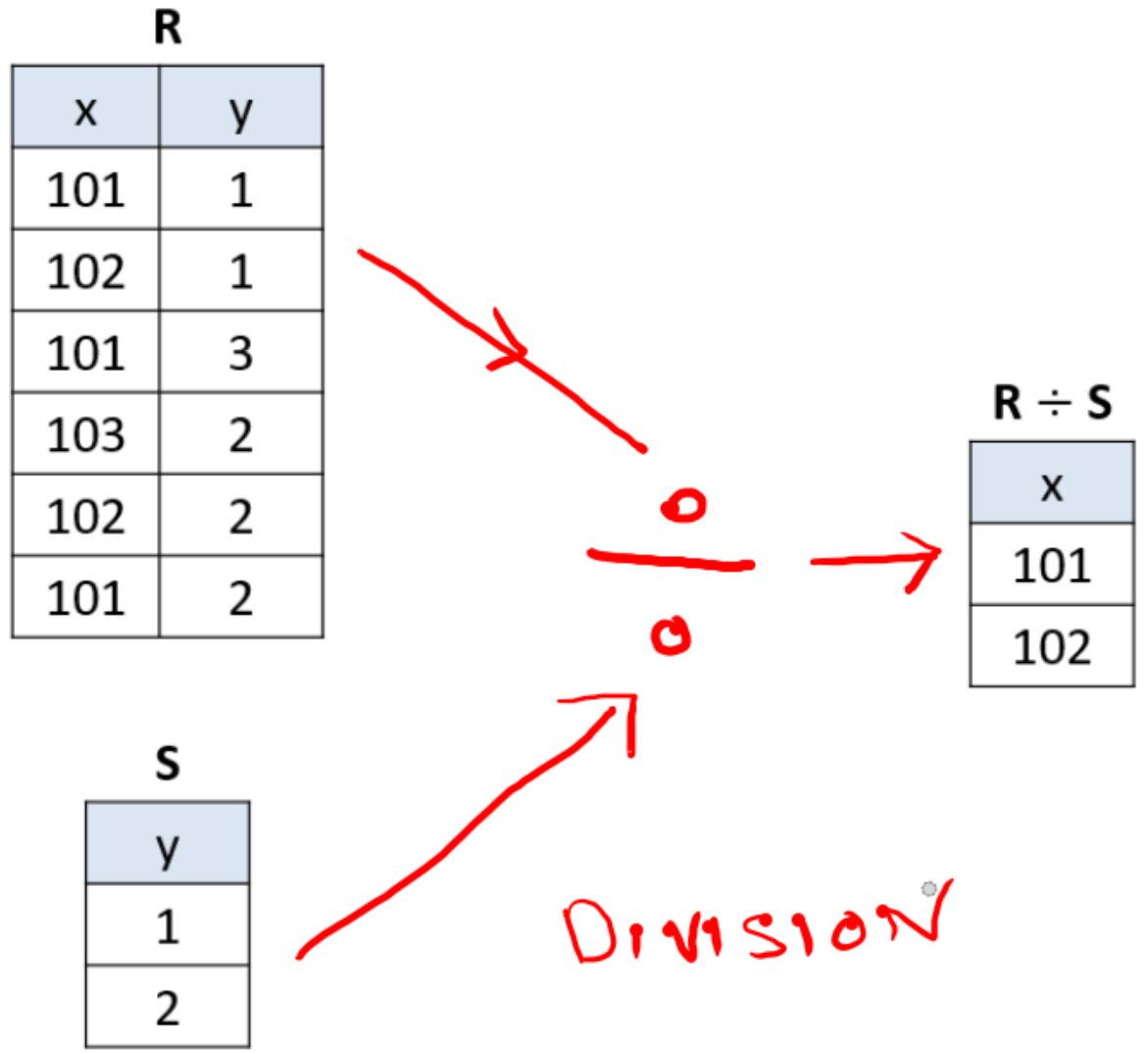
Division operation

- Following are example queries that require division
  - **`Supply-Parts`** database: Suppliers that supply all parts
  - Company database: List employees who work on all projects controlled by dno=4.
- The division is typically required when you want to find entities that are interacting with all entities of a given set.
- It is not supported in SQL implementations .. can be represented using other operations … bit complex

Division- definition

Given two relations; r(x,y) , s(y)

r ÷ s gives all distinct values of x from r that are associated with all values of y in s.



Division Examples

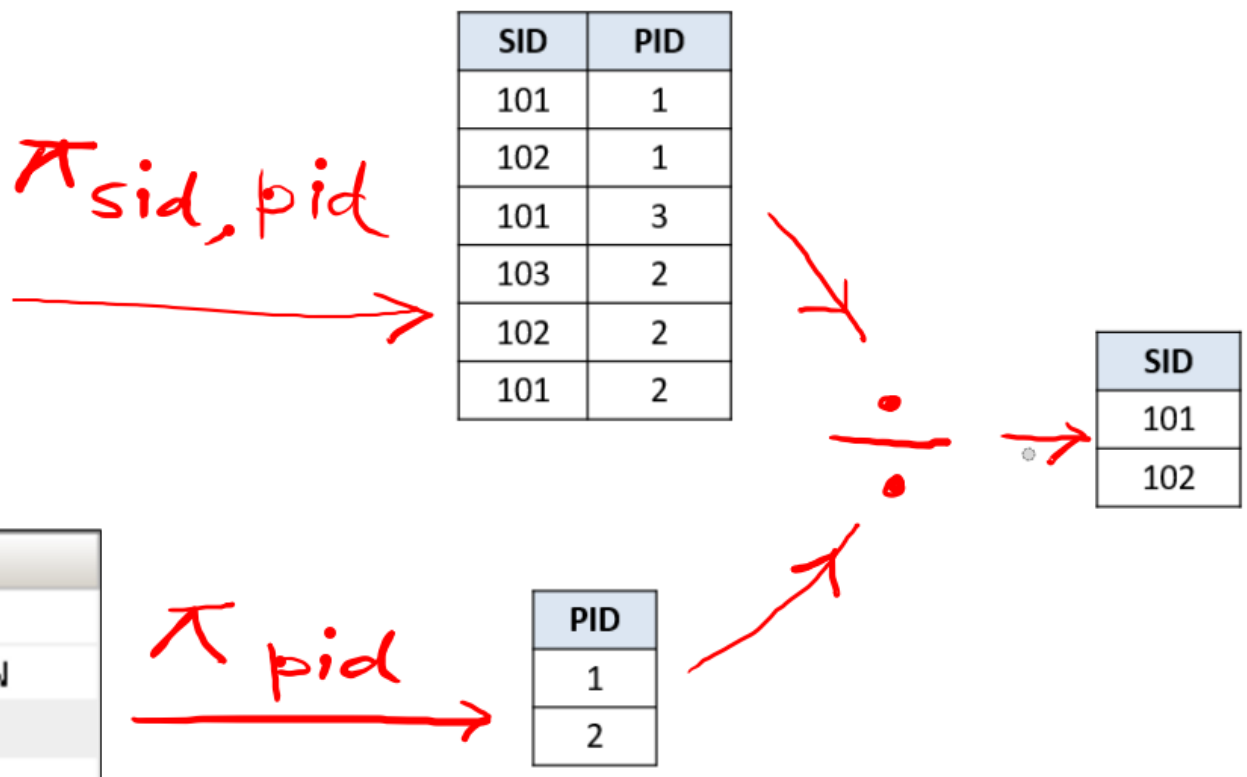
- Note that original relations may not be division compatible and required to brought down. As shown here! $\Pi_{sid,pid}(Supplies) \div \Pi_{pid}(Parts)$

supplies

sid	pid	cost
101	1	100
102	1	120
101	3	160
103	2	210
102	2	220
102	3	150
102	4	400
102	5	500

parts

pid	pname	color
1	PART-1	RED
2	PART-2	GREEN
3	PART-3	RED
4	PART-4	BLUE
5	PART-5	GREEN



SID	PID
101	1
102	1
101	3
103	2
102	2
101	2

PID
1
2

SID
101
102

Division Examples

List employees who work on all projects controlled by dno=4

- PNOs controlled by dno = 4
  $p4 \leftarrow \pi_{PNO}(\sigma_{DNO=4}(PROJECTS))$
- Have ENO, PNO project of WORKS on relation–
  $r1 \leftarrow \pi_{ENO,\ PNO}(WORKS\_ON)$

- ENO of employees works on PNOs in p4: **r1 div p4**

pno	pname	plocation	dno
1	ProductX	Bangalore	5
2	ProductY	Sigapore	5
3	ProductZ	Houston	5
10	Computerization	London	4
20	Reorganization	Houston	1
30	SentAnalysis	London	4

eno	pno	hours
101	1	32.5
101	2	7.5
104	3	40
103	1	20
103	2	20
102	2	10
102	3	10
102	10	10
102	20	10
102	1	32.5
108	30	30
108	10	10
107	10	35
107	30	5
106	30	20
106	20	15
105	20	

pno	pname	plocation	dno
1	ProductX	Bangalore	5
2	ProductY	Sigapore	5
3	ProductZ	Houston	5
10	Computerization	London	4
20	Reorganization	Houston	1
30	SentAnalysis	London	4

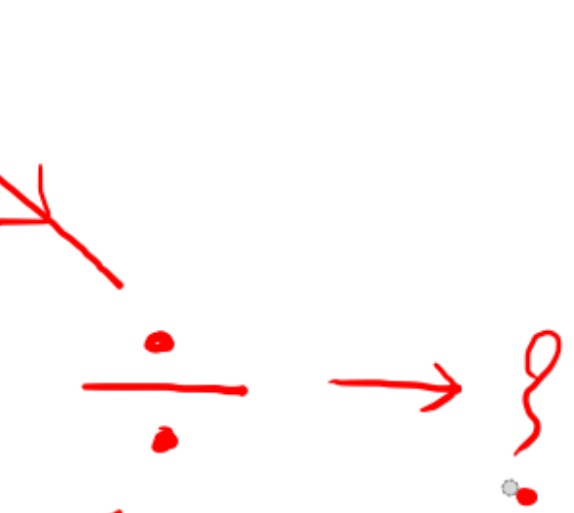
$\pi_{PNO}(\sigma_{DNO=4}(P))$

pno smallint
10
30

eno	pno	hours
101	1	32.5
101	2	7.5
104	3	40
103	1	20
103	2	20
102	2	10
102	3	10
102	10	10
102	20	10
102	1	32.5
108	30	30
108	10	10
107	10	35
107	30	5
106	30	20
106	20	15
105	20	

$\pi_{ENO,\ PNO}$

ssn numeric(9,0)	pno smallint
101	2
101	3
101	10
101	20
101	1
102	30
102	20
103	30
103	10
104	3
105	1
105	2
106	10
106	30
107	1
107	2
108	20



Division – computation

R

x	y
101	1
102	1
101	3
103	2
102	2
101	2

S

y
1
2

- One of the two approaches for computing **R div S** is as follows:
- Suppose XR is a set having the distinct values of X from R, that is $\Pi_x(R)$
- For division, we need to determine $x\ in\ XR$ that are associated with all elements $y\ in\ S$
- Let us assume all x are associated with all y; if so $XR\ \times S$ would be equal to R. Isn't it?
- If $x\ in\ XR \neq R$, then not all x are associated with all y
- Then $\Pi_x(XR - R)$ would be set having x that are not associated with all y, and $XR - \Pi_x(XR - R)$ would have x that are associated with all y

Division – computation

- Computation of r DIV s
- Note the compatibility of R and S
- Compute following and observe the result –

$r1 \leftarrow \Pi_x(r) \times s$ = ?

$r2 \leftarrow r1 - r$ = ?

$r2x \leftarrow \Pi_x(r2)$ = ?

$r3 \leftarrow \Pi_x(r) - r2x$ = ?

R

x	y
101	1
102	1
101	3
103	2
102	2
101	2

S

y
1
2

Division – computation

- Compute following and observe the result –

$r1 \leftarrow \Pi_x(r) \times s$ = ?

$r2 \leftarrow r1 - r$ = ?

$r2x \leftarrow \Pi_x(r2)$ = ?

$r3 \leftarrow \Pi_x(r) - r2x$ = ?

All possible combinations
r1 ← π_x(R) x S
x values with "incomplete combinations",
r2x ← π_x(r1-R)
and result - **π_x(R)-r2x**

$$\pi_x(R) - \pi_x((\pi_x(R) \times S) - R)$$

SQL Solution

R(x,y) DIV S(y) be expressed as

```
SELECT DISTINCT x FROM R
WHERE x NOT IN (
  SELECT x FROM (
     ( All possible; i.e. S x π_x(R) )
        MINUS
     ( Actual R )
  )
);
```

SELECT x that are
NOT IN
[CROSS – ACTUAL]

R

x	y
101	1
102	1
101	3
103	2
102	2
101	2

S

y
1
2

Division in SQL

supplies1

SID	PID
101	1
102	1
101	3
103	2
102	2
101	2

part1

PID
1
2

"Suppliers that supply all parts"

```
SELECT sid FROM Supplies1
WHERE sid NOT IN (
  SELECT sid FROM (
     ( All possible sid, pid combinations)
       MINUS
     ( Actual sid, pid pairs from Supplies1 )
);
```

$\pi_{SID}(SUPPLIES1) - \pi_{SID}((\pi_{SID}(SUPPLIES1) \times PARTS1) - SUPPLIES1)$

Division in SQL

supplies1

SID	PID
101	1
102	1
101	3
103	2
102	2
101	2

part1

PID
1
2

"Suppliers that supply all parts"

```
SELECT sid FROM supplies1
WHERE sid not in (
SELECT sid FROM (
  (SELECT * FROM (select distinct sid from supplies1) as sp
                  cross join part1
  EXCEPT
  (SELECT * FROM supplies1)
 ) AS r
);
```

$\pi_{SID}(SUPPLIES1) - \pi_{SID}((\pi_{SID}(SUPPLIES1) \times PARTS1) - SUPPLIES1)$

Division in SQL

"Suppliers that supply all parts"

```
SELECT * FROM suppliers
WHERE sid not in (
SELECT sid FROM (
  (SELECT sid, pid FROM (select pid from parts) as p cross
  join (select distinct sid from supplies) as sp)
  EXCEPT
  (SELECT sid, pid FROM supplies)
 ) AS r
);
```

supplies

sid	pid	cost
101	1	100
102	1	120
101	3	160
103	2	210
102	2	220
102	3	150
102	4	400
102	5	500

sid [PK] integer	sname character vary	city character varying
101	ABC	Mumbai
102	PQR	Delhi
103	XYZ	Ahmedabad

parts

pid	pname	color
1	PART-1	RED
2	PART-2	GREEN
3	PART-3	RED
4	PART-4	BLUE
5	PART-5	GREEN

$\Pi_{sid,pid}(Supplies)\ div\ \pi_{pid}(Parts)$

Division in SQL (example #2)

List employees who work on all projects controlled by dno=4

```
SELECT * FROM EMPLOYEE
WHERE eno NOT IN (
    SELECT eno FROM (
        ( All possible eno, pno (of dno=4) combinations)
        MINUS
        ( Actual eno, pno pairs from WORKS_ON )
);
```

Division in SQL (example #2)

pno	pname	plocation	dno
1	ProductX	Bangalore	5
2	ProductY	Sigapore	5
3	ProductZ	Houston	5
10	Computerization	London	4
20	Reorganization	Houston	1
30	SentAnalysis	London	4

List employees who work on all projects controlled by dno=4

```
SELECT * FROM employee AS e
WHERE eno NOT IN (
    SELECT eno FROM (
    (SELECT eno, pno FROM (select pno from project where dno=4) as p
            cross join (select distinct eno from works_on) as w)
    EXCEPT
    (SELECT eno, pno FROM works_on)
) AS r
);
```

eno	pno	hours
101	1	32.5
101	2	7.5
104	3	40
103	1	20
103	2	20
102	2	10
102	3	10
102	10	10
102	20	10
102	1	32.5
108	30	30
108	10	10
107	10	35
107	30	5
106	30	20
106	20	15
105	20	

Division in SQL Solution (Another Strategy)

R(x,y) DIV S(y) be expressed as

```
SELECT DISTINCT x FROM R
WHERE empty-set (
    ( all y, i.e. S )
      MINUS
    ( y that are associate with the x)
);
```

R

x	y
101	1
102	1
101	3
103	2
102	2
101	2

S

y
1
2

Division in SQL Solution (Another Strategy)

"Suppliers that supply all parts"

```
SELECT suppliers
WHERE empty-set (
    ( All Parts )
     MINUS
    ( Parts Supplied by the Supplier )
);
```

Division in SQL Solution (Another Strategy)

"Suppliers that supply all parts"

```
SELECT * FROM suppliers as s
WHERE NOT EXISTS (
    ( SELECT p.pid FROM parts as p )
     EXCEPT
    (SELECT sp.pid FROM supplies sp WHERE sp.sid = s.sid )
);
```

For division correlated query seems simpler to write but may expensive to execute

supplies

sid	pid	cost
101	1	100
102	1	120
101	3	160
103	2	210
102	2	220
102	3	150
102	4	400
102	5	500

parts

pid	pname	color
1	PART-1	RED
2	PART-2	GREEN
3	PART-3	RED
4	PART-4	BLUE
5	PART-5	GREEN

suppliers

sid [PK] integer	sname character var	city character varying
101	ABC	Mumbai
102	PQR	Delhi
103	XYZ	Ahmedabad

Division Example #2 (using s

pno	pname	plocation	dno
1	ProductX	Bangalore	5
2	ProductY	Sigapore	5
3	ProductZ	Houston	5
10	Computerization	London	4
20	Reorganization	Houston	1
30	SentAnalysis	London	4

eno	pno	hours
101	1	32.5
101	2	7.5
104	3	40
103	1	20
103	2	20
102	2	10
102	3	10
102	10	10
102	20	10
102	1	32.5
108	30	30
108	10	10
107	10	35
107	30	5
106	30	20
106	20	15
105	20	

List employees who work on all projects controlled by dno=4

```
SELECT employee
WHERE empty-set (
      ( all PNOs controlled by dno=4, i.e. p4 )
         MINUS
      ( PNOs on which the employee works)
);
```

Using Strategy#2

List employees who work on all projects controlled by dno=4

```
SELECT * FROM employee AS e
WHERE NOT EXISTS (
  (SELECT pno FROM project WHERE dno = 4)
   EXCEPT
  (SELECT pno FROM works_on AS w WHERE w.eno = e.eno)
);
```

Division Example#3

Students taken all courses that PMJ offered from academic year 2007-08 to 2011-12.

$$r1 \leftarrow \sigma_{iname='PMJ'}(instructor)$$

$$r2 \leftarrow \sigma_{acadyr>=2007 \text{ and } acadyr<=2011}(offers)$$

$$r3 \leftarrow r1 * r2 * registers$$

$$r4 \leftarrow \Pi_{sid,course,acadyear,semester}(r3)$$

$$r5 \leftarrow \Pi_{course,acadyear,semester}(r3)$$

$$result \leftarrow r4 \text{ div } r5$$

Division Example#3

Students taken all courses that PMJ offered from academic year 2007-08 to 2011-12.

[Using Strategy#1]

SELECT Students

WHERE sid NOT IN (

(All possible combination of sid, cno, yr, sem for PMJ and during specified acad-years)

MINUS

(actual combination of sid, cno, yr, sem in registers for PMJ and during specified acad-years)

);

Division Example#3

Students taken all courses that PMJ offered from academic year 2007-08 to 2011-12.

[Using Strategy#1]

```
SELECT * FROM student AS s
WHERE studentid NOT IN (
   SELECT studentid FROM (
      SELECT studentid, courseno, acadyear, semester from
         ((select courseno, acadyear, semester FROM offers NATURAL JOIN instructor
          WHERE instructorname = 'P M Jat' AND acadyear >= 2007 AND  acadyear <= 2011) as co
            CROSS JOIN (select distinct studentid from registers) as sr)
          EXCEPT
            (SELECT studentid, courseno, acadyear, semester FROM
               registers WHERE acadyear >= 2007 AND acadyear <= 2011)
   ) as r
);
```

Division Example#3

Students taken all courses that PMJ offered from academic year 2007-08 to 2011-12.

[Using Strategy#2]

SELECT Students

WHERE empty-set (

(All courses by PMJ and during specified acad-years)

MINUS

(Courses taken by the StudID during specified acad-years)

);

Division Example#3

Students taken all courses that PMJ offered from academic year 2007-08 to 2011-12.

[Using Strategy#2]

```
SELECT * FROM student AS s
WHERE NOT EXISTS (
        (SELECT courseno, acadyear, semester FROM offers NATURAL JOIN instructor WHERE instructorname = 'P M Jat' AND acadyear >= 2007 AND acadyear <= 2011)
        EXCEPT
        (SELECT courseno, acadyear, semester FROM registers AS r WHERE acadyear >= 2007 AND acadyear <= 2011 AND r.studentid=s.studentid)
);
```

More queries requiring DIVISION

- Retrieve the names of employees, who work on all the projects that ‘John Smith’ works
- List supplier who supply all ‘Red’ Parts
- List all customers who bought all items for category=3