In the Lecture Series Introduction to Database Systems

Relational Algebra

Presented by **Stéphane Bressan**

Relational Algebra

- An operational query language
- A query in relational algebra is
  A formula of relations and operators
- Formulae can be translated into SQL queries
- SQL queries can be optimized using relational algebra

Operations (Operators)

- Operations on a single relation
  - selection σ, projection π
- Usual set operations (relations are sets):
  - union $\cup$, intersection $\cap$, and difference $-$ (non-symmetric)
- Operations combining two or more relations
  - Cartesian product $\times$, join $\bowtie$, equi-join $\bowtie_E$ and natural join $\bowtie_n$
- A renaming operation ρ
- A division operation / (for self-study)

Example: employee

name	salary	eNumber
Clark	150000	1006
Gates	5000000	1005
Jones	50000	1001
Peter	45000	1002
Phillips	25000	1004
Rowe	35000	1003
Warnock	500000	1007

Example: plane

maker	mNumber
Airbus	A310
Airbus	A320
Airbus	A330
Airbus	A340
Boeing	B727
Boeing	B747
Boeing	B757
MD	DC10
MD	DC9

Example: canFly

eNumber	mNumber
1001	B727
1001	B747
1001	DC10
1002	A320
1002	A340
1002	B757
1002	DC9
1003	A310
1003	DC9
1003	DC9

Example: assigned

eNumber	date	fNumber
1001	Nov 1	100
1001	Oct 31	100
1002	Nov 1	100
1002	Oct 31	100
1003	Oct 31	100
1003	Oct 31	337
1004	Oct 31	337
1005	Oct 31	337
1006	Nov 1	991
1006	Oct 31	337

Projection

Keeps vertical slices of a relation according to a list L of attributes (i.e. *a list of columns*) of the relation R:

$$\pi_L(R) = \{t \mid \exists\ t_1$$
$$(t_1 \in R \wedge_{A \in L}\ t.A = t_1.A)\}$$

Projection (Example)

$\pi_{\text{eNumber, fNumber}}$**(assigned)**

eNumber	date	fNumber
1001	Nov 1	100
1001	Oct 31	100
1002	Nov 1	100
1002	Oct 31	100
1003	Oct 31	100
1003	Oct 31	337
1004	Oct 31	337
1005	Oct 31	337
1006	Nov 1	991
1006	Oct 31	337

Projection (Result)

$$\pi_{\text{eNumber, fNumber}}(\text{assigned})$$

eNumber	fNumber
1001	100
1002	100
1003	100
1003	337
1004	337
1005	337
1006	991
1006	337

Projection (SQL)

```
SELECT DISTINCT eNumber, fNumber
FROM assigned
```

Selection

Selects the tuples of a relation verifying a condition c:

$$\sigma_c(R) = \{t \mid t \in R \wedge c\}$$

c is any Boolean expression ($\wedge$, $\vee$ $\neg$) involving t ($<$, $=$, $>$, $\neq$, $\leq$, $\geq$)

Selection (Example)

$$\sigma_{\ salary<100000}(\mathbf{employee})$$

name	salary	eNumber
Clark	150000	1006
Gates	5000000	1005
Jones	50000	1001
Peter	45000	1002
Phillips	25000	1004
Rowe	35000	1003
Warnock	500000	1007

Selection (Result)

$$\sigma_{\ salary<100000}\text{(employee)}$$

name	salary	eNumber
Jones	50000	1001
Peter	45000	1002
Phillips	25000	1004
Rowe	35000	1003

Selection (SQL)

```
SELECT *
FROM employee
WHERE salary < 100000
```

Selection (Example)

$$\sigma_{salary>100000 \wedge \neg(name='Gates')}(employee)$$

name	salary	eNumber
Clark	150000	1006
Gates	5000000	1005
Jones	50000	1001
Peter	45000	1002
Phillips	25000	1004
Rowe	35000	1003
Warnock	500000	1007

Selection (Result)

$$\sigma_{\text{salary}>100000\ \wedge\neg(\text{name}=\text{'Gates'})}(\text{employee})$$

name	salary	eNumber
Clark	150000	1006
Warnock	500000	1007

Selection (SQL)

```
SELECT *
FROM employee
WHERE salary > 100000
AND name <> 'Gates'
```

Remark: Composability

The result of a query is a relation

$\sigma_{salary< 50000}$ (employee)

$\pi_{name,\ salary}(\sigma_{salary< 50000}$ (employee))

Remark: Commutativity

$$\pi_{\text{name, salary}}(\sigma_{\text{salary} < 50000} \text{ (employee))}$$

$$\sigma_{\text{salary} < 50000}(\pi_{\text{name, salary}} \text{ (employee))}$$

Can we always do this?

Remark: SQL

$\pi_{\text{name, salary}}(\sigma_{\text{salary} < 50000}\ (\text{employee}))$

SELECT DISTINCT name, salary

FROM employee

WHERE salary < 50000

Projection

Selection

Union, Intersection, Set-difference

- $R_1 \cup R_2 = \{ t \mid t \in R_1 \vee t \in R_2 \}$
- $R_1 \cap R_2 = \{ t \mid t \in R_1 \wedge t \in R_2 \}$
- $R_1 — R_2 = \{ t \mid t \in R_1 \text{ and } \neg(t \in R_2) \}$

The relations R_1 and R_2 must be

union compatible:

- Same number of attributes
- Corresponding attributes have the same type (but not necessarily the same name)

Union (Example)

$plane_1 \cup plane_2$

$plane_1$

maker	mNumber
Airbus	A310
Airbus	A320
Airbus	A330
Boeing	B747
Boeing	B757

$plane_2$

maker	mNumber
Airbus	A330
Airbus	A340
Boeing	B727
Boeing	B747
MD	DC10
MD	DC9

Union (Result)

$plane_1 \cup plane_2$

maker	mNumber
Airbus	A310
Airbus	A320
Airbus	A330
Airbus	A340
Boeing	B727
Boeing	B747
Boeing	B757
MD	DC10
MD	DC9

Union (SQL)

```
SELECT *
FROM plane1
UNION
SELECT *
FROM plane2
```

What about duplicates?

Union

Find all the information about students in the computer science department or in the information systems department.

```
SELECT *
FROM student T
WHERE T.department='CS'
UNION
SELECT *
FROM student T
WHERE T.department='IS';
```

Intersection (Example)

$$\text{plane}_1 \cap \text{plane}_2$$

Plane$_1$

maker	mNumber
Airbus	A310
Airbus	A320
Airbus	A330
Boeing	B747
Boeing	B757

Plane$_2$

maker	mNumber
Airbus	A330
Airbus	A340
Boeing	B727
Boeing	B747
MD	DC10
MD	DC9

Intersection (Result)

$$\text{plane}_1 \cap \text{plane}_2$$

maker	mNumber
Airbus	A330
Boeing	B747

Intersection (SQL)

```
SELECT *
FROM plane1
INTERSECT
SELECT *
FROM plane2
```

What about duplicates?

Intersection

Find the emails of students in the computer science department owning a book with ISBN14 ‘978-0684801520’.

```
SELECT T1.email
FROM student T1
WHERE T1.department='CS'
INTERSECT
SELECT T2.owner AS email
FROM copy T2
WHERE T2.book='978-0684801520';
```

Difference (Example)

$plane_1 — plane_2$

$Plane_1$

maker	mNumber
Airbus	A310
Airbus	A320
Airbus	A330
Boeing	B747
Boeing	B757

$Plane_2$

maker	mNumber
Airbus	A330
Airbus	A340
Boeing	B727
Boeing	B747
MD	DC10
MD	DC9

Difference (Result)

$\text{plane}_1 - \text{plane}_2$

maker	mNumber
Airbus	A310
Airbus	A320
Boeing	B757

Difference (SQL)

```
SELECT *
FROM plane1
MINUS (EXCEPT)
SELECT *
FROM plane2
```

What about duplicates?

(Non-Symmetric) Difference

Find the mails of students in the computer science department except those owning a book with ISBN14 ‘978-0684801520’.

```
SELECT T1.email
FROM student T1
WHERE T1.department='CS'
EXCEPT
SELECT T2.owner AS email
FROM copy T2
WHERE T2.book='978-0684801520';
```

Oracle uses ``MINUS’’

Cartesian Product

Combines the tuples of two relations in all possible ways

$$R1 \times R2 = \{t \mid \exists\, t_1\, \exists\, t_2 \\ (t_1 \in R_1 \wedge t_2 \in R_2 \\ \wedge_{A \in R1}\ t.A = t_1.A \\ \wedge_{A \in R2}\ t.A = t_2.A)\}$$

Cartesian Product (Example)

canFly × plane

eNumber	mNumber
1001	B727
1001	B747
1001	DC10
1002	A320
1002	A340
1002	B757
1002	DC9
1003	A310
1003	DC9
1003	DC10

maker	mNumber
Airbus	A310
Airbus	A320
Airbus	A330
Airbus	A340
Boeing	B727
Boeing	B747
Boeing	B757
MD	DC10
MD	DC9

Cartesian Product (Result)

canFly × plane

90 tuples

eNumber	mNumber	maker	mNumber
1001	B727	Airbus	A310
1001	B727	Airbus	A320
1001	B727	Airbus	A330
1001	B727	Airbus	A340
1001	B727	Boeing	B727
1001	B727	Boeing	B747
1001	B727	Boeing	B757
1001	B727	MD	DC10
1001	B727	MD	DC9
1001	B747	Airbus	A310
1001	B747	Airbus	A320
1001	B747	Airbus	A330
1001	B747	Airbus	A340
1001	B747	Boeing	B727
1001	B747	Boeing	B747
1001	B747	Boeing	B757
1001	B747	MD	DC10
1001	B747	MD	DC9
1001	B727	Airbus	A310
1001	B727	Airbus	A320
…	…	…	…

Cartesian Product (SQL)

```
SELECT *
FROM canFly, plane
```

Cartesian Product

```
SELECT *
FROM student T1 CROSS JOIN book T2;
```

This is always equivalent to:

```
SELECT *
FROM student T1, book T2;
```

Join (θ-Join)

Combines the tuples of two relations that verify a condition

$$R_1 \bowtie_c R_2 = \{t \mid \exists\ t_1\ \exists\ t_2$$

$$(t_1 \in R_1 \wedge t_2 \in R_2 \wedge c$$

$$\wedge_{A\ \in\ R1}\ \ t.A = t_1.A$$

$$\wedge_{A\ \in\ R2}\ \ t.A = t2.A)\}$$

$$= \sigma_c\ (R_1 \times R_2)$$

Join (θ-Join) (Example)

$$canFly \bowtie_{canFly.mNnumber=plane.mNumber} plane$$

eNumber	mNumber
1001	B727
1001	B747
1001	DC10
1002	A320
1002	A340
1002	B757
1002	DC9
1003	A310
1003	DC9
1003	DC10

maker	mNumber
Airbus	A310
Airbus	A320
Airbus	A330
Airbus	A340
Boeing	B727
Boeing	B747
Boeing	B757
MD	DC10
MD	DC9

Join (Result)

canFly $\bowtie_{\text{canFly.mNnumber=plane.mNumber}}$ plane

eNumber	mNumber	maker	mNumber
1001	B727	Boeing	B727
1001	B747	Boeing	B747
1001	DC10	MD	DC10
1002	A320	Airbus	A320
1002	A340	Airbus	A340
1002	B757	Boeing	B757
1002	DC9	MD	DC9
1003	A310	Airbus	A310
1003	DC9	MD	DC9
1003	DC10	MD	DC10

Join (SQL)

```
SELECT *
FROM canFly c, plane p
WHERE c.mNumber = p.mNumber
```

Inner Join

Find the emails of students owning a book with ISBN14 ‘978-0262033848’

```
SELECT T1.email
FROM student T1 INNER JOIN copy T2
ON T2.owner=T1.email
WHERE T2.book='978-0684801520';
```

Why would one want to do that?

Natural Join

- Combines two relations on a condition composed only of equalities of attributes with the same name in the first and second relation
- Projects only one of the redundant attributes (since they are equal)

$$R_1 \bowtie_N R_2$$

Natural Join (Example)

canFly $\bowtie_N$ plane

eNumber	mNumber
1001	B727
1001	B747
1001	DC10
1002	A320
1002	A340
1002	B757
1002	DC9
1003	A310
1003	DC9
1003	DC10

maker	mNumber
Airbus	A310
Airbus	A320
Airbus	A330
Airbus	A340
Boeing	B727
Boeing	B747
Boeing	B757
MD	DC10
MD	DC9

Natural Join (Result)

canFly $\bowtie_N$ plane

eNumber	mNumber	maker
1001	B727	Boeing
1001	B747	Boeing
1001	DC10	MD
1002	A320	Airbus
1002	A340	Airbus
1002	B757	Boeing
1002	DC9	MD
1003	A310	Airbus
1003	DC9	MD

Natural Join

Find the emails of students lending a a book and the ISBN of the book

```
SELECT T1.owner, T2.book
FROM copy T1 NATURAL JOIN loan T2;
```

(incidentally this query is equivalent to the following because of the foreign key constraint)

```
SELECT T2.owner, T2.book
FROM loan T2;
```

Renaming

Renaming a relation or its attributes:

$$\rho(R'(N_1 \rightarrow N'_1, \ldots, N_n \rightarrow N'_n), R)$$

The new relation R' has the same instance as R, but its schema has attribute N'_i instead of attribute N_i

Renaming (Example)

$\rho(\text{staff}(\text{salary} \rightarrow \text{wages}), \text{employee})$

employee

name	salary	eNumber
Clark	150000	1006
Gates	5000000	1005
Jones	50000	1001
Peter	45000	1002
Phillips	25000	1004
Rowe	35000	1003
Warnock	500000	1007

Renaming (Result)

$\rho(\text{staff}(\text{salary} \rightarrow \text{wages}), \text{employee})$

staff

name	wages	eNumber
Clark	150000	1006
Gates	5000000	1005
Jones	50000	1001
Peter	45000	1002
Phillips	25000	1004
Rowe	35000	1003
Warnock	500000	1007

Renaming (SQL)

```
SELECT name, salary AS wages, eNumber
FROM employee
```

Example

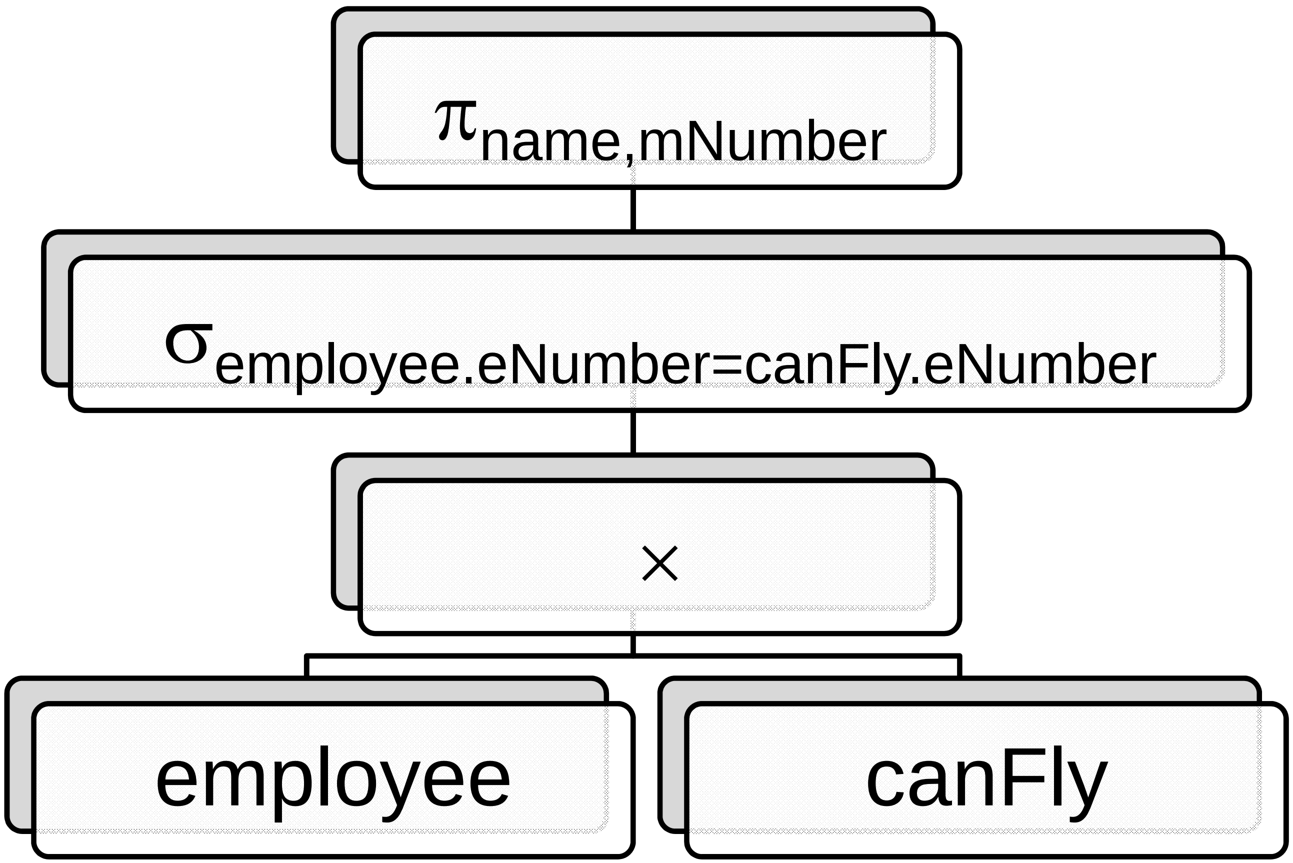
Find for each employee, her name and the model numbers of the planes she can fly

name	salary	eNumber
Clark	150000	1006
Gates	5000000	1005
Jones	50000	1001
Peter	45000	1002
Phillips	25000	1004
Rowe	35000	1003
Warnock	500000	1007

eNumber	mNumber
1001	B727
1001	B747
1001	DC10
1002	A320
1002	A340
1002	B757
1002	DC9
1003	A310
1003	DC9

Example

Find for each employee, her name and the model numbers of the planes she can fly



Remark: Project-Select-Join

Project-Select-Join (PSJ) queries correspond to simple SQL queries:

SELECT name, mNumber ← Projection

FROM employee, canFly ← Cartesian Product

WHERE employee.eNumber=canFly.eNumber

↑ Selection

Example

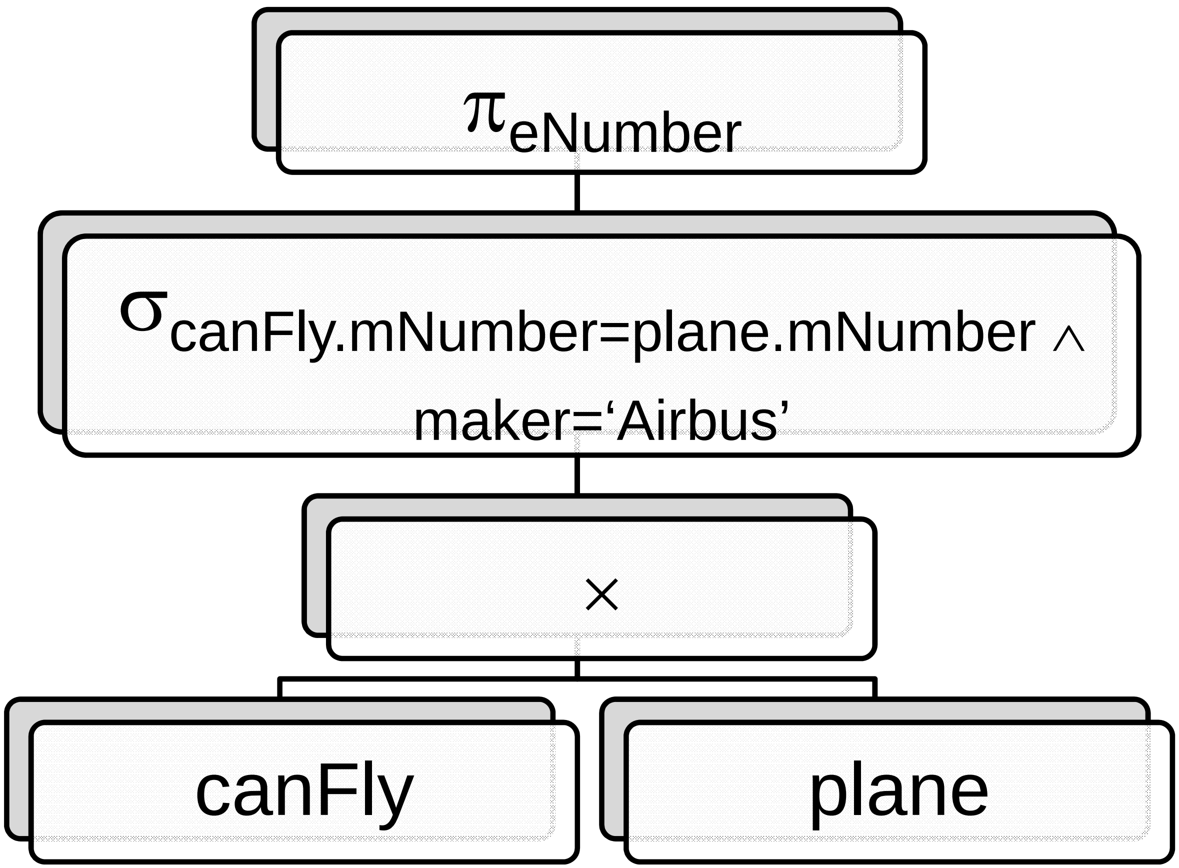
Find the employee numbers of employees who can fly Airbus planes

eNumber	mNumber
1001	B727
1001	B747
1001	DC10
1002	A320
1002	A340
1002	B757
1002	DC9
1003	A310
1003	DC9
1003	DC10

maker	mNumber
Airbus	A310
Airbus	A320
Airbus	A330
Airbus	A340
Boeing	B727
Boeing	B747
Boeing	B757
MD	DC10
MD	DC9

Example

Find the employee numbers of employees who can fly Airbus planes



Remark: Rewriting Algebra Expressions

- $\pi_{eNumber}\ (\sigma_{canFly.mNumber=plane.mNumber\ \wedge\ maker='Airbus'}\ (canFly \times plane))$

Which one is more efficient?

- $\pi_{eNumber}\ (\sigma_{canFly.mNumber=plane.mNumber}\ (canFly \times \sigma_{maker='Airbus'}(plane)))$

- $\pi_{eNumber}\ ((canFly \bowtie_{canFly.mNumber=plane.mNumber}\ \sigma_{maker='Airbus'}(plane)))$

- $\pi_{eNumber}\ (canFly \bowtie_{canFly.mNumber=plane.mNumber\ \wedge\ maker='Airbus'}\ plane)$

Division

- Consider two relations, R_1 with two attributes x and y, and R_2 with one attribute y
- R_1 / R_2 is the set of all x values such that for every y value in B, there is a tuple (x,y) in A
- Find all x which are related to every y in B.
  - Find the employees who can fly all MD planes
  - Find the students who have read all books by "Charles Dickens"
  - Find the actors who have acted in all movies by "Angelina Jolie"

Division

canFly

eNumber	mNumber
1001	B727
1001	B747
1001	DC10
1002	A320
1002	A340
1002	B757
1002	DC9
1003	A310
1003	DC9
1003	DC10

p1

mNumber
DC9

canFly/p1

eNumber
1002
1003

p2

mNumber
A320
A340

canFly/p2

eNumber
1002

Division

- Compute all possible combinations of column x of R_1 and R_2.

$$(\pi_x(R_1) \times R_2)$$

- Then remove those rows that exist in R_1

$$(\pi_x(R_1) \times R_2) - R_1$$

- Keep only the first column of the result. These are the ***disqualified*** values

$$\pi_x((\pi_x(R_1) \times R_2) - R_1)$$

- R_1/R_2 is the column x of R_1 **except** the disqualified values

$$R_1 / R_2 = \pi_x(R_1) - \pi_x((\pi_x(R_1) \times R_2) - R_1)$$

Example

Find the employment numbers of employees who can fly **all** MD planes

eNumber	mNumber
1001	B727
1001	B747
1001	DC10
1002	A320
1002	A340
1002	B757
1002	DC9
1003	A310
1003	DC9
1003	DC10

maker	mNumber
Airbus	A310
Airbus	A320
Airbus	A330
Airbus	A340
Boeing	B727
Boeing	B747
Boeing	B757
MD	DC10
MD	DC9

Example

Find the employment numbers of employees who can fly **all** MD planes

eNumber	mNumber
1001	B727
1001	B747
1001	DC10
1002	A320
1002	A340
1002	B757
1002	DC9
1003	A310
1003	DC9
1003	DC10

allMDPlanes:

$\pi_{mNumber}(\sigma_{maker='MD'}\ (plane))$

mNumber
DC10
DC9

canFly / allMDPlanes = ?

Division (Example)

Find the employment numbers of employees who can fly **all** MD planes

canFly / allMDPlanes =

canFly / $\pi_{\text{mNumber}}(\sigma_{\text{maker='MD'}}(\text{plane}))$ =

$$\pi_{\text{eNumber}}(\text{canFly}) - \pi_{\text{eNumber}}((\pi_{\text{eNumber}}(\text{canFly}) \times \pi_{\text{mNumber}}(\sigma_{\text{Maker='MD'}}(\text{plane}))) - \text{canFly})$$

Example: step-by-step

$$\pi_{eNumber}(canFly) \times \pi_{mNumber}(\sigma_{Maker='MD'}(plane))$$

eNumber	mNumber
1001	DC9
1001	DC10
1002	DC9
1002	DC10
1003	DC9
1003	DC10

Example: step-by-step

$$(\pi_{eNumber}(canFly) \times \pi_{mNumber}(\sigma_{Maker='MD'}(plane))) - canFly$$

eNumber	mNumber
1001	DC9
1002	DC10

Example: step-by-step

$$\pi_{eNumber}((\pi_{eNumber}(canFly) \times \pi_{mNumber}(\sigma_{Maker='MD'}(plane))) - canFly)$$

eNumber
1001
1002

Example: step-by-step

$$\pi_{eNumber}(canFly) - \pi_{eNumber}((\pi_{eNumber}(canFly) \times \pi_{mNumber}(\sigma_{Maker='MD'}(plane))) - canFly)$$

eNumber
1003

Division (SQL)

```
SELECT DISTINCT c1.eNumber
FROM canFly c1
WHERE NOT EXISTS (
   SELECT c.eNumber, p.mNumber
   FROM canFly c, plane p
   WHERE  p.maker='MD'  AND
          c1.eNumber=c.eNumber
          EXCEPT
          SELECT enumber, mNumber
          FROM canFly)
```

Division (SQL)

```
SELECT DISTINCT c1.eNumber
FROM canFly c1
WHERE NOT EXISTS (
   SELECT *
   FROM plane p
   WHERE p.maker='MD' AND NOT EXISTS(
         SELECT *
        FROM canFly c2
         WHERE c1.eNumber=c2.eNumber
                AND p.mNumber=c2.mNumber))
```

Credits

The content of this lecture is based on chapter 4 of the book "Introduction to database Systems"
By
S. Bressan and B. Catania, McGraw Hill publisher

Clipart and media are licensed from Microsoft Office Online Clipart and Media