Understanding the division operator in Relational Algebra

First lets take a quick look at how the operator works. Say we have two relations, R with schema: (A, B) and S with schema: (B). An instance of them could be:

A	B
a	1
a	2
a	7
b	1

B
1
7
2

If we name P and Q these instances of R and S respectively, the result of $P \div Q$ is:

A
a

What happened here? The division operator took all the values from attribute A that were in a relation (in R) with **all** the values of the B attribute in S. We were able to perform the operation because the last attribute of R had the same name as the attribute on S. In general, the last attributes of the first operator **must** have the same name as the attributes in the second operator and **must** be in the same order.

> Given two relations, $R1$ and $R2$ and their schema: $(A_1, \ldots, A_n, B_1, \ldots, B_m)$ and $(B_1, \ldots, B_m)$ resp., $R1 \div R2$ retrieves a relation $R3$ with schema $(A_1, \ldots, A_n)$ where all the sub-tuples $(A_1, \ldots, A_n)$ of $R1$ were in a relation with every tuple of $R2$.

Advanced example and explanation

Consider the following instances of the relations $R1$ and $R2$ defined before:

A	B	C	D	E
x_1	y_2	z_2	w_a	w_b
x_1	y_2	z_2	w_z	w_z
x_1	y_2	z_2	w_ϕ	w_z
x_1	y_2	z_2	w_b	w_q
x_1	y_2	z_2	w_w	w_u
x_1	y_2	z_2	w_b	w_b
x_1	y_2	z_2	w_x	w_h
x_1	y_2	z_2	w_q	w_a
x_2	y_5	z_1	w_a	w_b
x_2	y_4	z_1	w_a	w_b
x_3	y_7	z_3	w_a	w_x
x_3	y_7	z_3	w_a	w_b
x_3	y_7	z_3	w_x	w_h
x_3	y_7	z_3	w_b	w_q
x_3	y_7	z_3	w_w	w_u
x_3	y_7	z_3	w_z	w_z
x_3	y_1	z_1	w_a	w_b
x_3	y_7	z_3	w_b	w_b
x_3	y_7	z_3	w_ϕ	w_z

D	E
w_a	w_b
w_z	w_z
w_ϕ	w_z
w_b	w_q
w_w	w_u
w_b	w_b
w_x	w_h

Let's name X and Y the previous instances of $R1$ and $R2$ respectively.

Now, I will show a really easy way to understand the operator. It's nothing formal, just a way to see things. It consists on taking the table X and group it by the attributes (A, B, C), just like you do in SQL when using the **GROUP BY** statement. The groups are then represented by the value of their respective (A, B, C) attributes. Now we assign a set to each one of these groups, which contains all the pairs of values of the attributes (D, E) that are in a relation with the values that represent the group. Then we can check if the tuples of Y are included on each of these sets, if they are, then the values that represent that grouping are added to the solution, otherwise they are not. This is a bit confusing, but really easy to understand with the an example.

We start by identifying the rows of the table that are equal two to two on their (A, B, C) attributes. On table to the right we colored the identified groups.

A	B	C	D	E
x_1	y_2	z_2	w_a	w_b
x_1	y_2	z_2	w_z	w_z
x_1	y_2	z_2	w_ϕ	w_z
x_1	y_2	z_2	w_b	w_q
x_1	y_2	z_2	w_w	w_u
x_1	y_2	z_2	w_b	w_b
x_1	y_2	z_2	w_x	w_h
x_1	y_2	z_2	w_q	w_a
x_2	y_5	z_1	w_a	w_b
x_2	y_4	z_1	w_a	w_b
x_3	y_7	z_3	w_a	w_x
x_3	y_7	z_3	w_a	w_b
x_3	y_7	z_3	w_x	w_h
x_3	y_7	z_3	w_b	w_q
x_3	y_7	z_3	w_w	w_u
x_3	y_7	z_3	w_z	w_z
x_3	y_1	z_1	w_a	w_b
x_3	y_7	z_3	w_b	w_b
x_3	y_7	z_3	w_ϕ	w_z

If split the colored table by color we obtain five sub-tables:

A	B	C	D	E
x_1	y_2	z_2	w_a	w_b
x_1	y_2	z_2	w_z	w_z
x_1	y_2	z_2	w_ϕ	w_z
x_1	y_2	z_2	w_b	w_q
x_1	y_2	z_2	w_w	w_u
x_1	y_2	z_2	w_b	w_b
x_1	y_2	z_2	w_x	w_h
x_1	y_2	z_2	w_q	w_a

A	B	C	D	E
x_2	y_5	z_1	w_a	w_b

A	B	C	D	E
x_2	y_4	z_1	w_a	w_b

A	B	C	D	E
x_3	y_7	z_3	w_a	w_x
x_3	y_7	z_3	w_a	w_b
x_3	y_7	z_3	w_x	w_h
x_3	y_7	z_3	w_b	w_q
x_3	y_7	z_3	w_w	w_u
x_3	y_7	z_3	w_z	w_z
x_3	y_7	z_3	w_b	w_b
x_3	y_7	z_3	w_ϕ	w_z

A	B	C	D	E
x_3	y_1	z_1	w_a	w_b

Now let's see each of those tables like this:

- $(x_1, y_2, z_2) \rightarrow \{(w_a, w_b), (w_z, w_z), (w_\phi, w_z), (w_b, w_q), (w_w, w_u), (w_b, w_b), (w_x, w_h), (w_q, w_a)\}$

- $(x_2, y_5, z_1) \rightarrow \{(w_a, w_b)\}$

- $(x_2, y_4, z_1) \rightarrow \{(w_a, w_b)\}$

- $(x_3, y_7, z_3) \rightarrow \{(w_a, w_x), (w_a, w_b), (w_x, w_h), (w_b, w_q), (w_w, w_u), (w_z, w_z), (w_b, w_b), (w_\phi, w_z)\}$

- $(x_3, y_1, z_1) \rightarrow \{(w_a, w_b)\}$

Also let's see the table Y as a set:

- $\{(w_a, w_b), (w_z, w_z), (w_\phi, w_z), (w_b, w_q), (w_w, w_u), (w_b, w_b), (w_x, w_h)\}$

This is a clear representation of our data and we are ready to calculate the result of $X \div Y$. To simplify notation let's define a "function" in our universe, called **setOf()** that retrieves the set of tuples of what we pass to it as a parameter. Let's say, if we pass the parameter Y, it retrieves the set of it's tuples; if we pass three values x, y, z, it retrieves one of the sets defined before.

It's really easy to check that:

- $setOf(Y) \subseteq setOf(x_1, y_2, z_2)$

- $setOf(Y) \nsubseteq setOf(x_2, y_5, z_1)$

- $setOf(Y) \nsubseteq setOf(x_2, y_4, z_1)$

- $setOf(Y) \subseteq setOf(x_3, y_7, z_3)$

- $setOf(Y) \nsubseteq setOf(x_3, y_1, z_1)$

And the result of $X \div Y$ is:

A	B	C
x_1	y_2	z_2
x_3	y_7	z_3

I think this is a really good way to understand what the division operator does in relational algebra. This operator is the most confusing at first for some people, specially when using it with operands that have more than one attribute each.

Simple query example

Let's see a quick example of how to use it to write a query. Let's say we have a simplified version of Blockbuster's database and some relations of it's relational schema are:

CLIENTS(clientID , name, lastname)
MOVIES(movieID , title , description , year , duration , language)
RENTALS(clientID , movieID , date)

We want a relational algebra query that retrieves the ids of the clients that rented all the movies of the year 2010.

First of all we can find the ids of all the movies of the year 2010:

$A = \prod_{movieID}(\sigma_{year=2010}MOVIES)$

Now, let's project the relation RENTALS by clientID and movieID so we retrieve all the rentals ever done:

$B = \prod_{clientID,movieID} RENTALS$

Let's divide B by A! Note that the second row of B is the same that the only row of A. By doing this we are asking for all the clientIDs that are in a relation with all the movies of 2010. To finish the query we use the JOIN operator and project the result by name and lastname:

$Solution = \prod_{name,lastname}(CLIENTS * (B \div A))$