Division Operator in RA

How to approach a query in Relational Algebra that has keywords like "every" or "all"

You might have seen questions in Assignment 1 such as:
"Products supplied by every supplier"

To solve such queries in RA, we have a special operator known as **Division operator** denoted as "/". Division operator works exactly like division of two integers.

Let's try to understand division operator:

- Let R and S be relations with schemas $A_1, \ldots, A_n, B_1, \ldots, B_n$ and $B_1, \ldots, B_n$ respectively.
- The result of R/S is a relation T with
    - Schema $A_1, \ldots, A_n$ (attribute names in R but not in S)
    - Tuples **t** such that, for every tuple s of S, the tuple t||s (the concatenation of t and s) is in relation R
    - T contains the largest possible set of tuples s. t. $S \times T \subseteq R$
- Here is an analogy to integer division: –
    - For integers, A / B is: the largest int Q s.t. $Q \times B \leq A$
    - For relations, A / B is: the largest relation Q s.t. $Q \times B$

Consider following relations: R, S1, S2 and S3.
T1 is the result of division R/S1.
T2 is the result of division R/S2.
T3 is the result of division R/S3.

R

A	B	C
a1	b1	c1
a2	b1	c1
a1	b2	c1
a1	b2	c2
a2	b1	c2
a1	b2	c3
a1	b2	c4
a1	b1	c5

S1

C
c1

T1=R/S1

A	B
a1	b1
a2	b1
a1	b2

S2

C
c1
c2

T2=R/S2

A	B
a1	b2
a2	b1

S3

C
c1
c2
c3
c4

T3=R/S3

A	B
a1	b2

Some more examples of division:

R

A	B	C
a1	b1	c1
a2	b1	c1
a1	b2	c1
a1	b2	c2
a2	b1	c2
a1	b2	c3
a1	b2	c4
a1	b1	c5

S4

B	C
b1	c1

T4=R/S4

A
a1
a2

S5

B	C
b1	c1
b2	c1

T5=R/S5

A
a1

- Consider two relations A(x,y), B(y) and suppose we want to specify the query
  "Find all x's that are associated through A with all B's"
- This can be expressed as:
  - A/B = π_x (A) - π_x ((π_x (A) x B) - A)
  - Division operator is often useful when the query is about "every" or "all" (not limited for these keywords!)
  - Division operator doesn't extend the expressiveness of Relational Algebra since division can be expressed with existing operators in RA.

Don't worry if you are confused at this point.

Let's take an example and try to understand how division relates to
π_x (A) - π_x ((π_x (A) x B) - A)

Assume following two relations:

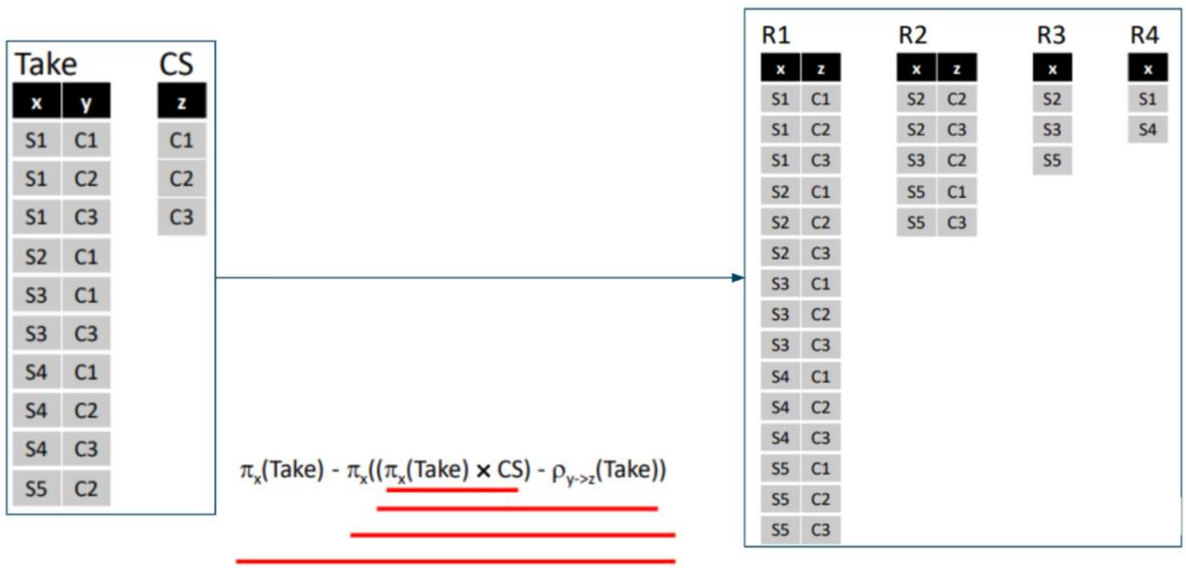
- Take(x,y) - "student x has taken course y",
- CS(z) - "z is a CS course"

We want **"All students who have taken all CS courses"**
We try to solve this problems in multiple steps.

- π_x(Take) x CS
  - Relation of all <student, CS course> pairs
- (π_x (Take) x CS)) - $\rho_{y\rightarrow z}$ (Take)
  - Relation of all <student, CS course> pairs that did NOT occur
- π_x ((π_x (Take) x CS) - $\rho_{y\rightarrow z}$ (Take))

- Relation of all <students> who have NOT taken all CS courses
- π_x (Take) - π_x ((π_x (Take) CS) - $\rho_{y \rightarrow z}$ (Take))
  - Relation of all <students> who have taken all CS courses

Pictorially,



You may use Division operator in your assignment.
Although you can express such queries without use of division operator as well.
As long as your query produces correct results, either one of these options will fetch you points in Assignment.

More practice:

Consider the running example of our university database:

- To run this example, load the [Silberschatz - UniversityDB] database in RelaX calculator, and run following example.
- Trace the resulting RA tree to understand each component.

Query: Find out all students who have taken **all** Physics courses.

AllPhyCourses = $\pi_{\text{course\_id}}\ \sigma_{\text{dept\_name = 'Physics'}}$ course
TakesCourse = $\pi_{\text{ID, course\_id}}$ takes
StudentsTakingAllPhyCourses = TakesCourse / AllPhyCourses