Relational Algebra Masterclass

Given:

Student(<u>SID</u>, Sname, GPA)

Department(<u>DName</u>, Chair, Building, Room)

Course(<u>DName</u>, <u>CID</u>, CName, Hours)

Enrolled(<u>DName</u>, <u>CID</u>, <u>SID</u>)

where

- DName in Course is a foreign key referencing Department,
- DName,CID in Enrolled is a foreign key referencing Course,
- SID in Enrolled is a foreign key referencing Student, and
- primary keys are underlined.

and the relation states

Student

SID	Sname	GPA
11	Bush	3.0
12	Cruz	3.2
13	Clinton	3.9
22	Sanders	3.0
33	Trump	3.8

Enrolled

DName	CID	SID
CS	101	11
Math	101	11
CS	101	12
CS	101	22
Math	103	33
EE	102	33
CS	102	22

Department

DName	Chair	Building	Room
CS	Rubio	Ajax	100
Math	Carson	Acme	300
EE	Kasich	Ajax	200
Music	Costello	North	100

Course

DName	CID	CName	Hours
CS	101	Programming	4
CS	102	Algorithms	3
Math	101	Algebra	3
Math	103	Calculus	4
Music	104	Jazz	3
EE	102	Circuits	3

Show how the following relational algebra expression gives the names of all students enrolled in two or more courses.

$$\pi_{SName}(\pi_{SID}(\sigma_{DName \neq D \text{ OR } CID \neq C}(\rho_{(D,C,SID)}(Enrolled) * Enrolled)) * Student)$$

Work from the inside out and play close attention to parentheses and operator-operand binding.

Apply $\rho_{(D,C,SID)}(Enrolled)$, which creates a relation like *Enrolled* but with *Dname* and *CID* renamed:

D	C	SID
CS	101	11
Math	101	11
CS	101	12
CS	101	22
Math	103	33
EE	102	33
CS	102	22

Then apply $\rho_{(D,C,SID)}(Enrolled) * Enrolled$, which natural joins the relation created above with *Enrolled*:

D	C	SID	DName	CID
CS	101	11	CS	101
CS	101	11	Math	101
Math	101	11	CS	101
Math	101	11	Math	101
CS	101	12	CS	101
CS	101	22	CS	101
CS	101	22	CS	102
Math	103	33	Math	103
Math	103	33	EE	102
EE	102	33	Math	103
EE	102	33	EE	102
CS	102	22	CS	101
CS	102	22	CS	102

Then apply $\sigma_{DName \neq D \text{ OR } CID \neq C}(\rho_{(D,C,SID)}(Enrolled) * Enrolled)$, which selects from the previous result only the rows for which $Dname \neq D$ or $CID \neq C$:

> Tip: You can view selection as choosing tuples for inclusion, or choosing tuples for elimination by negating the θ condition. By DeMorgan's Law $\neg\theta$ is $Dname = D \wedge CID = C$.

D	C	SID	DName	CID
CS	101	11	Math	101
Math	101	11	CS	101
CS	101	22	CS	102
Math	103	33	EE	102
EE	102	33	Math	103
CS	102	22	CS	101

From that result we project the SID attribute by applying $\pi_{SID}(...)$, which gives us:

SID
11
22
33

From there you can easily see that we have the *SID*s of all the students enrolled in two or more courses, which we natrual join with *Student* so we can project the *SName*s for the final result.