Relational Algebra Operations

Anne Denton

Department of Computer Science
North Dakota State University

Outline

Table of Contents

The operations on relations produce relations

- The input and the output of all relational algebra operations are relations
- That means that arbitrary sequences of operations are possible
- For operations that are based on sets, those sets can also be the result of relational algebra operations $\Rightarrow$ nesting
- Much of the flexibility of SQL derives from this observations

Set theoretic operations

- Since relations are sets of tuples, any set theoretic operations can be applied
- Union, intersection, and set difference are available for relations
- Cartesian product central to capability of combining relations

Specific operations to databases

- Select (σ) selects rows
- Project (π) selects columns
- Join ($\bowtie$) conceptually is a combination of select, project and cartesian product

Consider the following two relations

Student

sid	sName
2	Black
5	Green
9	Red

Enroll

sid	dept	cid	grade
2	CSci	366	A
5	CSci	366	B
2	CSci	474	B

Select

- Selects a subset of tuples from the relation
- $\sigma_{<\text{selection condition}>}$(R)
- Selection condition is a Boolean expression
- Select is commutative

Result of $\sigma_{(sid=5)}$(Enroll):

sid	**dept**	**cid**	**grade**
5	CSci	366	B

Question 1

How many rows result from the following $\sigma_{\text{(dept='CSci') AND (cid=366)}}$(Enroll)?

1. 0
2. 1
3. 2
4. 3

Project

- Projects to a subset of attributes the relation
- $\pi_{<\text{attribute list}>}$(R)
- The relational algebra project eliminates duplicate rows, but the SQL implementation does not
- Project is not commutative, because attributes second project has to be a subset of those in first

Result of π_{cid}(Enroll):

cid
366
474

In a relational database all 3 entries would be preserved (366 would be listed twice)

Brainstorming question

Why would it make sense for relational DBMSs to preserve duplicate entries?

Table of Contents

Set-theoretic operations

- Relations are sets of tuples
- Anything operation can be used on sets can be used on relations
- Union: $R \bigcup S$ includes tuples that are in R or in S
- Intersection: $R \bigcap S$ includes tuples that are in R and in S
- Set Difference: $R - S$ includes tuples that are in R but not in S

Result of π_{sid}(Student)-π_{sid}(Enroll):

sid
9

Question 2

Assume that you want to select students that take both CSci 366 and CSci 474?
Could you use $\sigma_{(cid=474)\mathrm{AND}(cid=366)}$(Enroll)
to produce that result?

Question 3

Assume that you want to select students that take both CSci 366 and CSci 474?
Could you use $\sigma_{(cid=474)}(\text{Enroll}) \bigcap \sigma_{(cid=366)}(\text{Enroll})$
to produce that result?

Question 4

Assume that you want to select students that take both CSci 366 and CSci 474?
Could you use $\pi_{sid}\left(\sigma_{(cid=474)}(\text{Enroll})\right) \bigcap \pi_{sid}\left(\sigma_{(cid=366)}(\text{Enroll})\right)$ to produce that result?

Table of Contents

Cartesian product

- Used to combine attributes from multiple relations
- Standard set theoretic operation
- Combines each element from one with each element from another
- Usually a selection is necessary to extract the useful ones

Question 5

How many elements does the Cartesian product of the sets $\{1, 2, 3\}$ and $\{A, B\}$ have

1. 0
2. 2
3. 3
4. 6

Question 6

How many elements does the Cartesian product of the following two relations $\{(1, A), (2, A), (3, B)\}$ and $\{(A, x), (B, y)\}$ have?

1. 0
2. 2
3. 3
4. 6

From Cartesian product to Join

In a database setting, one would expect the combination of the following relations

1	A
2	A
3	B

and

A	x
B	y

to produce

1	A	x
2	A	x
3	B	y

⇒ Selecting the rows with matching "join attribute" is called a natural join. After removing the redundant column the result is called natural join

Table of Contents

Cartesian Product Example

Result of Student × Enroll:

Student. sid	Student. sName	Enroll. sid	Enroll. dept	Enroll. cid	Enroll. grade
2	Black	2	CSci	366	A
2	Black	5	CSci	366	B
2	Black	2	CSci	474	B
5	Green	2	CSci	366	A
5	Green	5	CSci	366	B
5	Green	2	CSci	474	B
9	Red	2	CSci	366	A
9	Red	5	CSci	366	B
9	Red	2	CSci	474	B

Making the Cartesian product useful

Superfluous rows removed by selection and columns by projection:

$\pi_{Student.sid,Student.sName,Enroll.dept,Enroll.cid,Enroll.grade}($
$\sigma_{(Student.sid=Enroll.sid)}(\text{Student} \times \text{Enroll}))$

Student. sid	Student. sName	Enroll. dept	Enroll. cid	Enroll. grade
2	Black	CSci	366	A
2	Black	CSci	474	B
5	Green	CSci	366	B

Result of selection and projection called a natural join:
Student * Enroll

Types of joins

- **Student** $\bowtie$ **Enroll**: Inner join (also called equi join) does selection, but not projection and expects that join attribute specified for both
- In DBMSs the inner join is actually safer than the natural join
- **Student * Enroll**: Natural join does selection and projection and assumes that join attribute has same name
- In both cases only tuples included, where join attribute appears on both sides: Student with sid=9 not included!

$\Rightarrow$ Motivates Outer join

Outer Join

- **Student** $=\bowtie$ **Enroll**:Left Outer Join includes unmatched tuples for first relation
- **Student** $\bowtie=$ **Enroll**: Right Outer Join includes unmatched tuples for second relation
- **Student** $=\bowtie=$ **Enroll**: Full Outer Join includes unmatched tuples for both relations

Full Outer Join

Student. sid	Student. sName	Enroll. sid	Enroll. dept	Enroll. cid	Enroll. grade
2	Black	2	CSci	366	A
2	Black	2	CSci	474	B
5	Green	5	CSci	366	B
9	Red	<null>	<null>	<null>	<null>

Question 7

Consider the following two relations

aid	atext
1	x
2	xx
4	y
6	yz

and

bid	aid
123	2
456	2
111	4
222	<null>
333	4

How many rows does the inner join have (aid is join attribute)

1. 4
2. 5
3. 6
4. 7

Question 8

Consider the following two relations

aid	atext
1	x
2	xx
4	y
6	yz

and

bid	aid
123	2
456	2
111	4
222	<null>
333	4

How many rows does the left outer join have (aid is join attribute)

1. 4
2. 5
3. 6
4. 7

Question 9

Consider the following two relations

aid	atext
1	x
2	xx
4	y
6	yz

and

bid	aid
123	2
456	2
111	4
222	<null>
333	4

How many rows does the right outer join have (aid is join attribute)

1. 4
2. 5
3. 6
4. 7

Question 10

Consider the following two relations

aid	atext
1	x
2	xx
4	y
6	yz

and

bid	aid
123	2
456	2
111	4
222	<null>
333	4

How many rows does the full outer join have (aid is join attribute)

1. 4
2. 5
3. 6
4. 7

Table of Contents

Examples where operations could be defined recursively

For example

- Employee has supervisor
- Webpage links to another

Join between a relation and itself possible but

- Every level separate join
- No recursive joining of all levels as single relational algebra command

Question 11

Consider the following relation

aid1	aid2
1	2
1	3
2	4
4	5

Assume that you join the table with itself, joining aid2 with aid1 each time. How many rows does an inner join produce after one join?

1. 0
2. 1
3. 2
4. 4
5. 6

Question 12

Consider the following relation

aid1	aid2
1	2
1	3
2	4
4	5

Assume that you join the table with itself, joining aid2 with aid1 each time. How many rows does an inner join produce after two joins?

1. 0
2. 1
3. 2
4. 4
5. 6

Question 13

Consider the following relation

aid1	aid2
1	2
1	3
2	4
4	5

Assume that you join the table with itself, joining aid2 with aid1 each time. How many rows does an inner join produce after three joins?

1. 0
2. 1
3. 2
4. 4
5. 6

Question 14

Consider the following relation

aid1	aid2
1	2
1	3
2	4
4	5

Assume that you join the table with itself, joining aid2 with aid1 each time. How many rows does a left outer join produce after one join?

1. 2
2. 4
3. 5
4. 6

Question 15

Consider the following relation

aid1	aid2
1	2
1	3
2	4
4	5

Assume that you join the table with itself, joining aid2 with aid1 each time. How many rows do left outer joins produce after two joins?

1. 2
2. 4
3. 5
4. 6

Question 16

Consider the following relation

aid1	aid2
1	2
1	3
2	4
4	1

Assume that you join the table with itself, joining aid2 with aid1 each time. How many rows do left outer joins produce after one join?

1. 2
2. 4
3. 5
4. 6

Question 17

Consider the following relation

aid1	aid2
1	2
1	3
2	4
4	1

Assume that you join the table with itself, joining aid2 with aid1 each time. How many rows do left outer joins produce after two joins?

1. 2
2. 4
3. 5
4. 6

Problems with cycles

- Without cycles, inner joins eventually terminate
  - For inner joins no rows are left
  - For outer joins the number of rows no longer increases
- With cycles
  - The number of rows may increase indefinitely

Question 18 (Multiple answers may be correct)

Which of the following scenarios may create indefinitely growing relations upon repeated joining?

1. Employees and their supervisors
2. Webpages linking to webpages
3. Publications citing earlier publications

Also: What are the respective data structures called?