Algebra and Join Minimization

John Clara, Danyang Zhang

Winter 2016.

Relational Algebra

Review on Relational Algebra

Basic Relational Algebra

- Selection: $\sigma_{a=C}R$
- Projection: $\Pi_a R$
- Rename: $\delta_{a_1 \to a_2} R$
- Aggregation: $\gamma_{a,Fn(b)\to b'} R$
- Binary operator: $\cup, \cap, \bowtie, \div, \times, -$

Example Schema

Taken from SQL Lab Assignment 1.

S :

sailor	sname	rating

B :

boat	sname	bname	day

R :

reservation	bname	color	rating

Example

List the sailors who have at least one reservation and only reserved red boats.

Example

List the sailors who have at least one reservation and only reserved red boats.

$$\Pi_{sname} R - \Pi_{sname}((\sigma_{color \neq' red'} B) \bowtie R)$$

Example

List the sailor name pairs who reserve the same boat.

Example

List the sailor name pairs who reserve the same boat.

$$\Pi_{sname1,sname2}(\sigma_{sname1<sname2}(\\ (\Pi_{sname1,bname}\delta_{sname\rightarrow sname1}R) \bowtie \\ (\Pi_{sname2,bname}\delta_{sname\rightarrow sname2}R))\\ )$$

Example

List the sailor names who reserve every red boat (assuming there exists red boats). Hint: use $\div$.

Example

List the sailor names who reserve every red boat (assuming there exists red boats).
$\forall$ corresponds to $\div$.

$$\Pi_{sname}(R \div \sigma_{color='red'} B)$$

Example

List the sailor names who reserve every red boat ~~(assuming there exists red boats)~~. Hint: use two $-$.

Example

SQL using NOT IN:

```
select sname from sailor
where sname NOT IN (
  select sname from sailor, boat b
  where b.color = red and sname NOT IN (
    select sname from reservation
    where bname = b.bname))
```

Relational algebra

$$\Pi_{sname} S - \Pi_{sname}(\\ \Pi_{sname,boat}(\sigma_{color='red'} B \bowtie S) - \\ \Pi_{sname,boat}(\sigma_{color='red'} B \bowtie R) \\ )$$

Join Minimization

Join Minimization

How to Optimize Queries

Basic Rules:

- Perform different mappings to reduce rows
- Answer variables cannot map to others
- Constants cannot map to others
- Everything else is fair game!

Example:

R	title	author
r1	t	“Bob”
r2	a	-
r3	a	b

ans	title
	t

Example 1

What are all the books by the person who wrote “Twilight”?

Example 1

What are all the books by the person who wrote “Twilight”?

```
SELECT b1.title
FROM Book b1, Book b2, Book b3
WHERE b1.author = b2.author AND
      b3.author = b2.author AND
      b3.title = "Twilight";
```

Example 1

What are all the books by the person who wrote “Twilight”?

```
SELECT b1.title
FROM Book b1, Book b2, Book b3
WHERE b1.author = b2.author AND
      b3.author = b2.author AND
      b3.title = "Twilight";
```

Book	title	author
b1	d	a
b2	-	a
b3	“Twilight”	a

answer	title
	d

Can we map first row $b1$ to any rows?

Example 1

What are all the books by the person who wrote “Twilight”?

Book	title	author
b1	d	a
b2	-	a
b3	“Twilight”	a

answer	title
	d

Map second $b2$ row to some row?

Example 1

What are all the books by the person who wrote “Twilight”?

Book	title	author
b1	d	a
b3	“Twilight”	a

answer	title
	d

Map $b3$ to some row?

Example 1

What are all the books by the person who wrote “Twilight”?

Book	title	author
b1	d	a
b3	“Twilight”	a

answer	title
	d

```
SELECT b1.title
FROM Book b1, Book b3
WHERE b1.author = b3.author AND
      b3.title = "Twilight";
```

Example 2

```
SELECT t1.A, t2.B, t4.C
FROM R t1, R t2, R t3, R t4, R t5
WHERE t3.A=t4.A AND
      t2.B=t3.B AND
      t1.C=t2.C AND
      t3.C=t5.C AND
      t3.A=t5.A;
```

Example 2

```
SELECT t1.A, t2.B, t4.C
FROM R t1, R t2, R t3, R t4, R t5
WHERE t3.A=t4.A AND
      t2.B=t3.B AND
      t1.C=t2.C AND
      t3.C=t5.C AND
      t3.A=t5.A;
```

R	A	B	C
t1	a	-	c1
t2	-	b	c1
t3	a1	b	c2
t4	a1	-	c
t5	a1	-	c2

answer	A	B	C
	a	b	c

Example 2

R	A	B	C
t1	a	-	c1
t2	-	b	c1
t3	a1	b	c2
t4	a1	-	c
t5	a1	-	c2

answer	A	B	C
	a	b	c

Can we reduce any rows?

Example 2

Reduce t5

R	A	B	C
t1	a	-	c1
t2	-	b	c1
t3	a1	b	c2
t4	a1	-	c

answer	A	B	C
	a	b	c

Example 2

Can reduce t2 to t3 or t3 to t2, considering t1 and t4?

R	A	B	C
t1	a	-	c1
t2	-	b	c1
t3	a1	b	-
t4	a1	-	c

answer	A	B	C
	a	b	c

How to Chase

The Chase in detail

- Repeat until no change
  - For each $X \rightarrow A$ in F do
    - For all rows t_1, t_2 in P such that $t_1(X) = t_2(X)$, $t_1(A) \neq t_2(A)$ do
      - if $t_1(A)$, $t_2(A)$ are non-answer variables then replace one by the other everywhere in P
      - if $t_1(A)$ is a non-answer variable and $t_2(A)$ is a wildcard, then replace $t_2(A)$ by $t_1(A)$ everywhere in P
      - If $t_1(A)$, $t_2(A)$ are wildcards, replace both with a new variable
      - if $t_1(A)$ is an answer variable and $t_2(A)$ is a variable or wildcard, then replace $t_2(A)$ by $t_1(A)$ everywhere in P
      - if $t_1(A)$ is constant, $t_2(A)$ is variable or wildcard, then replace $t_2(A)$ by $t_1(A)$ everywhere in P
      - if $t_1(A)$ is constant, $t_2(A)$ is constant then STOP and output $\varnothing$

118

Example 2

Dependencies: $F = \{AC \rightarrow B, B \rightarrow C, C \rightarrow A\}$

R	A	B	C
	a	-	c1
	-	b	c1
	a1	b	-
	a1	-	c

answer	A	B	C
	a	b	c

Example 2

Dependencies: $F = \{AC \rightarrow B, B \rightarrow C, C \rightarrow A\}$

Use $B \rightarrow C$

R	A	B	C
	a	-	c1
	-	b	c1
	a1	b	-
	a1	-	c

answer	A	B	C
	a	b	c

Example 2

Dependencies: $F = \{AC \rightarrow B, B \rightarrow C, C \rightarrow A\}$

Use $C \rightarrow A$

R	A	B	C
	a	-	c1
	-	b	c1
	a1	b	c1
	a1	-	c

answer	A	B	C
	a	b	c

Example 2

Dependencies: $F = \{AC \rightarrow B, B \rightarrow C, C \rightarrow A\}$

Eliminate rows

R	A	B	C
	a	-	c1
	a	b	c1
	a	b	c1
	a	-	c

answer	A	B	C
	a	b	c

Example 2

Dependencies: $F = \{AC \rightarrow B, B \rightarrow C, C \rightarrow A\}$

Can we use any Dependencies?

R	A	B	C
	a	b	c1
	a	-	c

answer	A	B	C
	a	b	c

Example 2

Dependencies: $F = \{AC \rightarrow B, B \rightarrow C, C \rightarrow A\}$

R	A	B	C
	a	b	-
	a	-	c

answer	A	B	C
	a	b	c

```
SELECT r1.A, r1.B, r2.C
FROM R r1, R r2
WHERE r1.a = r2.a;
```

Example 3

```
SELECT t1.A, t2.B, t4.C
FROM R t1, R t2, R t3, R t4
  WHERE t2.C=5 AND t3.A=t4.A AND
  t2.B=t3.B AND t1.C=t2.C AND
  t4.A=8;
```

with functional dependencies:

$$F = \{AC \rightarrow B, B \rightarrow C, C \rightarrow A\}.$$

Example 3

```
SELECT t1.A, t2.B, t4.C
FROM R t1, R t2, R t3, R t4
  WHERE t2.C=5 AND t3.A=t4.A AND
  t2.B=t3.B AND t1.C=t2.C AND
  t4.A=8;
```

R	A	B	C
t1	α	-	5
t2	-	β	5
t3	8	β	-
t4	8	-	γ

answer	A	B	C
	α	β	γ

Example 3

$$F = \{AC \rightarrow B, B \rightarrow C, C \rightarrow A\}.$$

R	A	B	C
t1	α	-	5
t2	-	β	5
t3	8	β	-
t4	8	-	γ

answer	A	B	C
	α	β	γ

After Chase

R	A	B	C
t1	$\alpha = 8$	-	5
t2	8	β	5
t3	8	β	5
t4	8	-	γ

answer	A	B	C
	α	β	γ

Example 3

After join minimization:

R	A	B	C
t2	8	β	5
t4	8	-	γ

answer	A	B	C
	$\alpha = 8$	β	γ

The final SQL query:

```
SELECT t2.A, t2.B, t4.C
FROM R t2, R t4
WHERE t2.A=8 AND t2.C=5 AND t4.A=8;
```

Example 4

Given the following pattern, minimize the pattern.

```
SELECT t1.A, s1.E FROM R t1, R t2, R t3, R t4
S s1, S s2 WHERE
  t1.B=t2.B AND t2.C=t3.C AND
  t3.A=t4.A AND t4.B=s2.B AND
  s2.D=s1.D;
```

R	A	B	C
t1	α	b1	-
t2	-	b1	c
t3	a	-	c
t4	a	b	-

S	B	D	E
s1	-	d	ε
s2	b	d	-

answer	A	E
	α	ε

Example 4

R	A	B	C
t1	α	b1	-
t2	-	b1	c
t3	a	-	c
t4	a	b	-

S	B	D	E
s1	**b'**	d	ε
s2	b	d	-

answer	A	E
	α	ε

Example 4

R	A	B	C
t1	α	b1	-
t2	-	b1	c
t3	a	-	c
t4	a	**b'**	-

S	B	D	E
s1	b'	d	ε

answer	A	E
	α	ε

Example 4

R	A	B	C
t1	α	b1	-
t2	-	b1	c
t3	a	**b”**	c
t4	a	b’	-

S	B	D	E
s1	b’	d	ε

answer	A	E
	α	ε

Example 4

R	A	B	C
t1	α	b1	-
t2	-	b1	c
t3	a	b”	c

S	B	D	E
s1	**b”**	d	ε

answer	A	E
	α	ε

Example 4

R	A	B	C
t1	α	b1	-
t2	-	b1	c

S	B	D	E
s1	**b1**	d	ε

answer	A	E
	α	ε

Example 4

R	A	B	C
t1	α	b1	-

S	B	D	E
s1	b1	-	ε

answer	A	E
	α	ε

Reference

1. "*Database Systems Concepts*" by Silberschatz, Korth and Sudarshan, 6th edition, McGraw-Hill.