ADC Assignment 3

Harshit Rakesh Shiroiya

October 5, 2021

1. ```
   select e1.*
   from   E1 e1
   where  not exists (select distinct row() from F)
   union
   select e2.*
   from   E2 e2
   where  exists (select distinct row() from F);
   ```

   Ans:

   $$\pi_{e_1.*}\ (\pi_{e1.*}\ (E_1)\ -\ \pi_{e_1.*}\ (E_1\ \times\ \pi_{()}(F))\ \cup\ \pi_{e_2.*}(E_2\ \times\ \pi_{()}(F))$$

2. ```
   select not exists(select 1
                     from   R r1, R r2
                     where  r1.x <> r2.x) as fewerThanTwo;
   ```

   Ans:

   $$\pi_{(t)}\ (\pi_{r1.*}\ (R_1)\ -\ \pi_{r1.*}\ (R_1\ \bowtie_{r1.x<>r2.x}\ R_2)\ \cup\ \pi_{(f)}(R1\ \bowtie_{r1.x<>r2.x} R_2))$$

3. ```
   select L1(r1,...,rn)
   from   R1 r1, ..., Rn rn
   where  C1(r1,...,rn) and
             [not] exists (select L2(s1,...,sm)
                           from   S1 s1,..., S1 sm
                           where  C2(s1,...,sm,r1,...,rn)
                           [union | intersect | except]
                           select L3(t1,...,tk)
                           from   T1 t1, ..., Tk tk
                           where  C3(t1,...,tk,r1,...,rn))
   ```
   ```

Ans:

(a) exists (... union ...)

Step:

```
select L1(r1,...,rn)
from   (R1 r1, ..., Rn rn) , (S1 s1,..., S1 sm)
where  C1(r1,...,rn) and C2(s1,...,sm,r1,...,rn)
union
select L1(r1,...,rn)
from   (R1 r1, ..., Rn rn) , (T1 t1, ..., Tk tk)
where  C1(r1,...,rn) and C3(t1,...,tk,r1,...,rn))
```

RA Expression:

$$\pi_{L1(r1,..,rn)}(\sigma_{C1(r1,...,rn)\ \wedge C2(s1,...,rm)}\ (R_1 \times ... \times R_n \times S_1 \times ... \times S_m))$$
$$\cup$$
$$\pi_{L1(r1,..,rn)}(\sigma_{C1(r1,...,rn)\ \wedge C3(t1,...,tk)}\ (R_1 \times ... \times R_n \times T_1 \times ... \times T_k))$$

(b) Bounding Intersect Clause and Translating Exists in Where Clause

```
Select L1(r1...rn)
from R1 r1,...Rn rn,
(Select L2(s1..sm)
from S1 s1..Sm sm
where C2(s1..sm,r1..rn)
Intersect
Select L3(t1..tk)
from T1 t1..Tk tk
where C3(t1,..tk,r1,..rn))q
where C1(r1..rn)
```

$$\pi_{L1(r1..rn)}\ (\sigma_{C1(r1..rn)}\ (R1.. \times Rn \times ((\pi_{L2(s1..sm)}\ (\sigma_{C2(s1..sm,r1..rn)}\ (S1.. \times Sm)))$$
$$\wedge\ (\pi_{L3(t1..tk)}\ (\sigma_{C3(t1..tk,r1..rn)}\ (T1.. \times Tk))))))$$

(c) Bounding Except Clause and Translating Exists in Where Clause

```
Select L1(r1...rn)
from R1 r1,...Rn rn,
(Select L2(s1..sm)
from S1 s1..Sm sm
where C2(s1..sm,r1..rn)
Except
Select L3(t1..tk)
from T1 t1..Tk tk
```

```
where C3(t1,..tk,r1,..rn))q
where C1(r1..rn)
```

$$\pi_{L1(r1..rn)}\ (\sigma_{C1(r1..rn)}\ (R1.. \times Rn \times ((\pi_{L2(s1..sm)}\ (\sigma_{C2(s1..sm,r1..rn)}\ (S1.. \times Sm)))\ -\ (\pi_{L3(t1..tk)}\ (\sigma_{C3(t1..tk,r1..rn)}\ (T1.. \times Tk))))))$$

(d) Translating Not Exists in Where Clause

```
Select q.* from
(Select L1(r1...rn)
from R1 r1,...Rn rn
where C1(r1..rn)
Except
(Select L1(r1...rn)
from R1 r1,...Rn rn,S1..Sm
where C1(r1..rn) and C2(s1..sm,r1..rn)
Union
Select L1(r1...rn)
from R1 r1,...Rn rn,T1..tk
where C1(r1..rn) and C3(t1,..tk,r1,..rn)))q
```

$$\pi_{L1(r1..rn)}\ (\pi_{L1(r1..rn)}\ (\sigma_{C1(r1..rn)}\ (R1.. \times Rn)) - ((\pi_{L1(r1..rn)}\ (\sigma_{C1(r1..rn)\ \wedge\ C2(s1..sm,r1..rn)}\ (R1.. \times Rn \times S1.. \times Sm))\ \cup\ (\pi_{L1(r1..rn)}\ (\sigma_{C1(r1..rn)\ \wedge\ C3(t1..tk,r1..rn)}\ (R1.. \times Rn \times T1.. \times Tk))))))$$

(e) Bounding Intersect Clause and Translating Not Exists in Where Clause

```
Select q.* from
(Select L1(r1...rn)
from R1 r1,...Rn rn
where C1(r1..rn)
Except
(Select L1(r1...rn)
from R1 r1,...Rn rn,(Select q2.* from
(Select L2(s1..sm)
from S1..Sm
where C2(s1..sm,r1..rn)
Intersect
Select L3(t1...tk)
from T1..tk
where C3(t1,..tk,r1,..rn))q2
where C1(r1..rn))))q
```

$\pi_{L1(r1..rn)}\ (\pi_{L1(r1..rn)}\ (\sigma_{C1(r1..rn)}\ (R1.. \times Rn)) - ((\pi_{L1(r1..rn)}\ (\sigma_{C1(r1..rn)}$
$(R1.. \times Rn \times (\pi_{L2(s1..sm)}(\pi_{L2(s1..sm)}\ (\sigma_{C2(s1..sm,r1..rn)}\ (S1.. \times Sm))\ \wedge\ (\pi_{L3(t1..tk)}$
$(\sigma_{C3(t1..tk,r1..rn)}\ (T1.. \times Tk))))))$

(f) Bounding Except Clause and Translating Not Exists in Where Clause

```
Select q.* from
(Select L1(r1...rn)
from R1 r1,...Rn rn
where C1(r1..rn)
Except
(Select L1(r1...rn)
from R1 r1,...Rn rn,(Select q2.* from
(Select L2(s1..sm)
from S1..Sm
where C2(s1..sm,r1..rn)
Except
Select L3(t1...tk)
from T1..tk
where C3(t1,..tk,r1,..rn))q2
where C1(r1..rn))))q
```

$\pi_{L1(r1..rn)}\ (\pi_{L1(r1..rn)}\ (\sigma_{C1(r1..rn)}\ (R1.. \times Rn)) - ((\pi_{L1(r1..rn)}\ (\sigma_{C1(r1..rn)}$
$(R1.. \times Rn \times (\pi_{L2(s1..sm)}(\pi_{L2(s1..sm)}\ (\sigma_{C2(s1..sm,r1..rn)}\ (S1.. \times Sm))\ -\ (\pi_{L3(t1..tk)}$
$(\sigma_{C3(t1..tk,r1..rn)}\ (T1.. \times Tk))))))$

4. Let R be a relation with schema (a,b,c) and let S be a relation with schema (d,e).Prove, from first principles7, the correctness of the following rewrite rule:

   Ans:

$$\pi_{a,d}\ (R \bowtie_{c=d} S) = \ \pi_{a,d}\ (\pi_{a,c}(R) \bowtie_{c=d}\ \ \pi_d(S))$$

$$\begin{array}{l} Let\ x, y \in \pi_{a,d}\ (R \bowtie_{c=d} S) \\ x \in a\ and\ y \in d \\ x \in R.a\ and\ y \in S.d \\ x, c\ \in R.a,\ R.c\ and\ y \in S.d \\ x \in \pi_{a,d}\ (\pi_{a,c}(R) \bowtie_{c=d}\ \ \pi_d(S)) \\ \pi_{a,d}\ (R \bowtie_{c=d} S) \subseteq\ \pi_{a,d}\ (\pi_{a,c}(R) \bowtie_{c=d}\ \ \pi_d(S)) \end{array}$$

$$\begin{array}{l} Let\ x, y\ \in\ \pi_{a,d}\ (\pi_{a,c}(R)) \\ x, c\ \in R.a,\ R.c\ and\ y \in S.d \\ x \in R.a\ and\ y \in S.d \\ x \in R\ and\ y \in S \\ x \in \pi_{a,d}(R \bowtie_{c=d} S) \\ \pi_{a,d}\ (\pi_{a,c}(R) \bowtie_{c=d}\ \ \pi_d(S)) \subseteq \pi_{a,d}\ (R \bowtie_{c=d} S) \end{array}$$

$$\pi_{a,d}\ (R \bowtie_{c=d} S) = \pi_{a,d}\ (\pi_{a,c}(R) \bowtie_{c=d}\ \ \pi_d(S))$$

5. Let R be a relation with schema (a,b,c) and let S be a relation with schema (d,e).Prove, from first principles7, the correctness of the following rewrite rule:

   $\pi_{a,d}\ (R \bowtie_{c=d} S)$
   $\pi_{a,d}(R \bowtie S)$
   $\pi_{a,c}(R \bowtie S)$
   $\pi_{a,c}(R)$
   $The\ rewrite\ rule\ enforced\ here\ is\ \pi_L(R \bowtie S) = \pi_L(R \bowtie S) = \pi_L(R)$
   $Because$

   $\pi_{a,d}\ (R \bowtie_{c=d} S)$
   $\pi_{a,d}\ (R \bowtie S)$
   $\pi_{a,d}\ (\pi_{a,c}(R) \bowtie \pi_d(S))$
   $\pi_{R.a,S.d}\ (\pi_{a,c}(R) \bowtie \pi_d(S))$
   $\pi_{R.a,R.c}\ (\pi_{a,c}(R) \bowtie \pi_d(S))\ (As\ R.c\ references\ S.d)$
   $\pi_{a,c}\ (\pi_{a,c}(R) \bowtie \pi_d(S))$
   $\pi_{a,c}(R)$

6. Optimize this RA SQL query and provide the optimized expression in standard RA notation. Specify at least three conceptually different rewrite rules that you used during the optimization.

   Ans:

   $\pi_{c.cname,c.headquarter}(C \bowtie_{c.cname=w.cname} (\sigma_{w.salary<55000}W)) \bowtie_{w.pid=p.pid} (\sigma_{p.city<>'Bloomington'}(P))$

   Step 1 *Pushing projections over Joins*

   $\pi_{c.cname,c.headquarter}(C \bowtie_{c.cname=w.cname} (\pi_{w.cname,w.pid} (\sigma_{w.salary<55000}W)) \bowtie_{w.pid=p.pid} (\pi_{p.pid} (\sigma_{p.city<>'Bloomington'}(\pi_{p.pid,p.city}(P)))$

   Step 2 : *Solving*

   $\pi_{c.cname,c.headquarter}(C \bowtie_{c.cname=w.cname} (\pi_{w.cname,w.pid} (\sigma_{w.salary<55000}W)) \bowtie_{w.pid=p.pid} (\pi_{p.pid} (\sigma_{p.city<>'Bloomington'}(P))))$

   Step 3 : *Introducing Semi Joins*

   $\pi_{c.cname,c.headquarter}(C \bowtie_{c.cname=w.cname} (\pi_{w.cname,w.pid} (\sigma_{w.salary<55000}W)) \ltimes (\pi_{p.pid}(\sigma_{p.city<>'Bloomington'}(P))))$

7. Let R be a relation with schema (a,b,c) and let S be a relation with schema (d,e).Prove, from first principles7, the correctness of the following rewrite rule:

   Ans:

$$\pi_{pid}(\pi_{pid}(\pi_{pid(A-B))\ cap}$$
$$\pi_{pid}(C - \pi_{pid,pname,city}(D \cap \pi_{p.*}(E - F) \cap \pi_{p.*}(E - G))))$$
$$where, A = \pi_{p.pid,s.skill}(P \times S)$$
$$B = \pi_{p.pid,s.skill}(P \bowtie_{pS.pid=p.pid} (pS) \bowtie_{pS.skill=s.skill} (S))$$
$$C = \pi_{p.*}(P)$$
$$D = \pi_{p.*,s1.skill,s2.skill}(P \times S1 \bowtie_{s1.skill \neq s2.skill} (S2))$$
$$E = \pi_{p.*,s1.skill,s2.skill}(P \times S1 \times S2)$$
$$F = \pi_{p.*,s1.skill,s2.skill}(P \bowtie_{pS.pid=p.pid} (pS) \bowtie_{pS.skill=s1.skill} (S1) \times S2)$$
$$G = \pi_{p.*,s1.skill,s2.skill}(P \bowtie_{pS.pid=p.pid} (pS) \bowtie_{pS.skill=s2.skill} (S2) \times S1)$$

   Step 1: Attribute Elimination

$$\pi_{pid}(\pi_{pid}(\pi_{pid(A-B))\ cap}$$
$$\pi_{pid}(C - \pi_{pid}(D \cap \pi_{pid}(E - F) \cap \pi_{pid}(E - G)))$$

   Step 2: Joins to Natural Joins

$$\pi_{pid}(\pi_{pid}(\pi_{pid(A-B))\ cap}$$
$$\pi_{pid}(C - \pi_{pid}(D \cap \pi_{pid}(E - F) \cap \pi_{pid}(E - G)))$$

   Step 3: Removing the projection

$$\pi_{pid}(\pi_{pid(A-B))\ cap}$$
$$\pi_{pid}(C - \pi_{pid}(D \cap \pi_{pid}(E - F) \cap \pi_{pid}(E - G)))$$

   Step 4: Pushing down projection over join

$$\pi_{pid}(\pi_{pid(A-B))\ cap}$$
$$\pi_{pid}(C - \pi_{pid}(D \cap \pi_{pid}(E - F) \cap \pi_{pid}(E - G)))$$

8. Optimize this RA SQL query and provide the optimized expression in standard RA notation. Specify at least three conceptually different rewrite rules that you used during the optimization.

   Ans:

   `Step 1: RA QUERY`

   $$\pi_{p.pid,p.pname}\ (\pi_{p.pid,p.pname}\ (\sigma_{p.pid=w.pid\ \wedge\ w.cname=c.cname\ \wedge\ c.city='Bloomington'}\ (P \times W \times C))\ \cap\ (\pi_{p.pid,p.pname}(P)\ -\ \pi_{p.pid,p.pname}(\sigma_{p.pid=k.pid1\ \wedge\ k.pid2=p1.pid\ \wedge\ p1.city='Chicago'}\ (P \times K \times P_1)))$$

   `Step 2: SEPERATING AND CONDITIONS`

   $$\pi_{p.pid,p.pname}\ (\pi_{p.pid,p.pname}\ (\sigma_{p.pid=w.pid}\ (\sigma_{w.cname=c.cname}(\ \sigma_{c.city='Bloomington'}\ (P \times W \times C))))\ \cap\ (\pi_{p.pid,p.pname}(P)\ -\ \pi_{p.pid,p.pname}(\sigma_{p.pid=k.pid1}(\ \sigma_{k.pid2=p1.pid}\ \sigma_{p1.city='Chicago'}(\ (P \times K \times P_1))))$$

   `Step 3: CONVERTING THE CROSS JOINS TO JOINS`

   $$\pi_{p.pid,p.pname}\ (\pi_{p.pid,p.pname}\ (P \bowtie W \bowtie (\ \sigma_{c.city='Bloomington'}\ C)\ \cap\ (\pi_{p.pid,p.pname}(P)\ -\ \pi_{p.pid,p.pname}(P \bowtie K \bowtie (\sigma_{p1.city='Chicago'}\ P_1)))$$

   `Step 4: USING SEMI-JOINS`

   $$\pi_{p.pid,p.pname}\ (\pi_{p.pid,p.pname}\ (P \bowtie \pi_{w.pid,w.cname}(W) \ltimes \sigma_{c.city='Bloomington'}(C))\ \cap\ (\pi_{p.pid,p.pname}(P\ -\ (\pi_{p.pid,p.pname}(P) \ltimes K \bowtie \pi_{pid}(\sigma_{p1.city='Chicago'}\ P_1)))$$

9. A $= \pi_{w.*}(\sigma_{w.salary<=70000}(W))$
   $B = \pi_{ps.pid}(\sigma_{ps.skill<=Programming}(PS))$
   $D = \pi_{ps.pid}(\sigma_{ps.skill='AI'})$

$\pi_{cname,headquarter}(\pi_{c.cname,c.headquarter}(C) \bowtie (\pi_{w.pid,w.cname}(W) -$
$(\pi_{cname,headquarter}(\pi_{w.pid,c.cname,c.headquarter}(A) \bowtie C -$
$\pi_{w.pid,c.cname,c.headquarter}(A) \bowtie C \bowtie_{w.pid=ps.pid} (B)) \cup$
$(\pi_{cname,headquarter}(\pi_{w.pid,c.cname,c.head}(A) \bowtie C -$
$\pi_{w.pid,c.cname,c.headquarter}(A) \bowtie C \bowtie_{w.pid=ps.pid} (D)))))))$

Step 1: Removing Unnecessary Projections

$\pi_{cname,headquarter}(C) \bowtie (\pi_{w.pid,w.cname}(W)) - \pi_{cname,headquarter}$
$(\pi_{w.pid,c.cname,c.headquarter}(A) \bowtie \pi_{c.cname}(C) - \pi_{w.pid,c.cname,c.headquarter}(A)$
$\bowtie C \bowtie_{w.pid=ps.pid} (B)) \cup (\pi_{cname,headquarter}(\pi_{w.pid,c.cname,c.head}(A)$
$\bowtie C - \pi_{w.pid,c.cname,c.headquarter}(A) \bowtie C \bowtie_{w.pid=ps.pid} (D)))))$

10. Optimize this RA SQL query and provide the optimizedexpression in standard RA notation.

    Ans:

    `Step 1: RA QUERY`

    $$\pi_{p.pid,f}\ (\pi_{p.pid,p.pname}(P)\ -\ \pi_{p.pid,p.pname}(\sigma_{hm1.mid=p.pid\ \wedge\ hm2.mid=p.pid\ \wedge\ hm1.eid<>hm2.eid}(P \times hm1 \times hm2))\ \cup\ \pi_{p.pid,t}(\sigma_{hm1.mid=p.pid\ \wedge hm2.mid=p.pid\ \wedge\ hm1.eid<>hm2.eid}(P \times hm1 \times hm2))$$

    `Step 2: SEPRATING ANDs`

    $$\pi_{p.pid,f}\ (\pi_{p.pid,p.pname}(P)\ -\ \pi_{p.pid,p.pname}(\sigma_{hm1.mid=p.pid}\ (\sigma_{hm2.mid=p.pid}(\sigma_{\ hm1.eid<>hm2.eid}(P \times hm1 \times hm2))))\ \cup\ \pi_{p.pid,t}(\sigma_{hm1.mid=p.pid}(\sigma_{hm2.mid=p.pid}(\sigma_{hm1.eid<>hm2.eid}(P \times hm1 \times hm2)))$$

    `Step 3: CONVERTING THE CROSS JOINS TO JOINS`

    $$\pi_{p.pid,f}\ (\pi_{p.pid,p.pname}(P)\ -\ \pi_{p.pid,p.pname}(hm1 \bowtie_{hm1.mid=p.pid} P \bowtie_{p.pid=hm2.mid\ \wedge\ hm1.eid<>hm2.eid}\ hm2)))\ \cup\ \pi_{p.pid,f}(hm1 \bowtie_{hm1.mid=p.pid} P\ \bowtie_{hm1.eid<>hm2.eid\ \ \wedge\ hm2.mid=p.pid} hm2)$$

    `Step 4: SIMPLFYING AND USING NATURAL JOINS`

    $$\pi_{p.pid,f}\ (\pi_{p.pid,p.pname}(P)\ -\ \pi_{p.pid,p.pname}(\pi_{mid,eid}\ (hm1) \bowtie \pi_{pid}(P)\ \bowtie_{hm1.eid<>hm2.eid} \pi_{mid,eid}(hm2)))\ \cup\ \pi_{pid,t}(hm1 \bowtie P \bowtie_{hm1.eid<>hm2.eid} hm2)$$