Fall B561 Assignment 3: RA Solutions
Name: Sricharaann Ramaswamy; Username: sriramas; UID: 2000855651

Collabrated with ATHULYA ANAND, SRI VARSHA CHELLAPILLA, SIDDHANT MESHRAM

1 Theoretical problems related to query translation and optimization

1. RA is,

$$\pi_{e1.*}(e1) - \pi_{e1.*}(e1 \times F)$$

Therefore RA expression is,

$$\pi_{q1.*}(q1) \cup -\pi_{e2.*}(e2 \times F)$$

2. RA is,

$$\pi_T - \pi_T(\sigma_{r1.x \neq r2.x}(r1 \times r2 \times T))$$

Therefore RA expression is,

$$\pi_Q(Q) \cup -\pi_F(\sigma_{r1.x<>r2.x}(F \times r1 \times r2))$$

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

```
select L1(r1,...,rn)
from   R1 r1, ..., Rn rn, S1 s1, ..., Sm 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:

$$\pi L1(r1,..,rn)(\sigma_{C1(r1,...,rn)\ \wedge C2(s1,...,rm)}\ (R1 \times ... \times R_n \times S_1 \times ... \times S_m))$$
$$\cup$$
$$\pi L1(r1,..,rn)(\sigma_{C1(r1,...,rn)\ \wedge C3(t1,...,tk)}\ (R1 \times ... \times R_n \times T_1 \times ... \times T_k))$$

(b) exists (... intersect ...)

```
select L1q(r1,...,rn)
from ( select r1.*, ..., rn.*
   from R1 r1, ..., Rn rn,S1 s1, ...,Sm sm
   where  C1(r1,...,rn) and C2(s1, ..., sm,r1,...,rn)
   intersect
   select r1.*, ...,rn.*
   from   R1 r1, ..., Rn rn, T1 t1, ..., Tk tk
   where  C1(r1,...,rn) and C3(t1,...,tk,r1, ...,rn)
) q
```

RA Expression:

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

(c) exists (... except ...)

```
select L1q(r1,...,rn)
from ( select r1.*, ..., rn.*
   from R1 r1, ..., Rn rn,S1 s1, ...,Sm sm
   where  C1(r1,...,rn) and C2(s1, ..., sm,r1,...,rn)
   except
   select r1.*, ...,rn.*
   from   R1 r1, ..., Rn rn, T1 t1, ..., Tk tk
   where  C1(r1,...,rn) and C3(t1,...,tk,r1, ...,rn)
) q
```

RA Expression:

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

(d) not exists (... union ...)
Step 1: NOT EXISTS to EXISTS

```
select L1q(r1,...,rn)
from ( select r1.*, ..., rn.*
   from R1 r1, ..., Rn rn
   where  C1(r1,...,rn)
   except
   select r1.*, ...,rn.*
   from   R1 r1, ..., Rn rn
   where  exists (select L2(s1, ...,sm)
      from S1 s1, ...,Sm, ...,sm)
      where C2(s1, ...,sm,r1, ...,rn)
      union
      select L3(t1, ..., tk)
      from T1 t1, ..., Tk tk
      where C3(t1, ...,tk,r1, ...,rn)
    )
) q
```

Step 2:

```
select L1q(r1,...,rn)
from ( select r1.*, ..., rn.*
   from R1 r1, ..., Rn rn
   where  C1(r1,...,rn)
   except
 ( select r1.*, ...,rn.*
   from   R1 r1, ..., Rn rn, S1 s1, ...,Sm sm
   where C2(s1, ...,sm,r1, ...,rn)
   union
   select r1.*,...,rn.*
   from R1 r1, ...,Rn rn,T1 t1, ..., Tk tk
   where C3(t1, ...,tk,r1, ...,rn)
  )
) q
```

RA Expression:

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

(e) not exists (... intersect ...)
Step 1: NOT EXISTS to EXISTS

```
select L1q(r1,...,rn)
from ( select r1.*, ..., rn.*
   from R1 r1, ..., Rn rn
   where  C1(r1,...,rn)
   except
   select r1.*, ...,rn.*
   from   R1 r1, ..., Rn rn
   where  exists (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
```

Step 2:

```
select L1q(r1,...,rn)
from ( select r1.*, ..., rn.*
   from R1 r1, ..., Rn rn
   where  C1(r1,...,rn)
   except
   select q1.*
   from ( select r1.*, ...,rn.*
      from   R1 r1, ..., Rn rn, S1 s1, ...,Sm sm
      where C2(s1, ...,sm,r1, ...,rn)
      intersect
      select r1.*,...,rn.*
      from R1 r1, ...,Rn rn,T1 t1, ..., Tk tk
      where C3(t1, ...,tk,r1, ...,rn)
  ) q1
) q
```

RA Expression:

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

(f) not exists (... except ...)
Step 1: NOT EXISTS to EXISTS

```
select L1q(r1,...,rn)
from ( select r1.*, ..., rn.*
   from R1 r1, ..., Rn rn
   where  C1(r1,...,rn)
   except
   select r1.*, ...,rn.*
   from   R1 r1, ..., Rn rn
   where  exists (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
```

Step 2:

```
select L1q(r1,...,rn)
from ( select r1.*, ..., rn.*
   from R1 r1, ..., Rn rn
   where  C1(r1,...,rn)
   except
   select q1.*
   from ( select r1.*, ...,rn.*
      from   R1 r1, ..., Rn rn, S1 s1, ...,Sm sm
      where C2(s1, ...,sm,r1, ...,rn)
      except
      select r1.*,...,rn.*
      from R1 r1, ...,Rn rn,T1 t1, ..., Tk tk
      where C3(t1, ...,tk,r1, ...,rn)
  ) q1
) q
```

RA Expression:

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

4. Rewriting RA,

$$\begin{aligned} LHS &= r|\exists R, \exists s(s\epsilon S \land s.d = r.c) \\ &= r|\exists S(a, c\epsilon R d\epsilon S \land s.d = r.c) \\ &= r|\exists S(a, c\epsilon R \land s.d = r.c) \\ &= \pi_{a,d}(\pi_{a,c}(R) \bowtie_{c=d} \pi_{d(s)} \\ &= RHS \end{aligned}$$

5. Rewriting RA,

$$\begin{aligned} \pi_{a,d}(R \bowtie_{c=d} S) &= \pi_{a,d}(\pi_{a,c}(R)) \bowtie_{c=d} \pi_d(S) \\ &= \pi_{a,c}(R \ltimes S) \end{aligned}$$

Reason: From the table R we know that, c is the foreign key referencing primary key d in S.

2 Translating Pure SQL queries to RA expressions and optimized RA expressions

6.

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

Step 1: Pushing down projections over joins

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

Step 2: Using semijoins

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

Step 3: Rewriting projection

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

7.

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

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.

$$\pi_{pid,pname}(\pi_{p.pid,pname}(\sigma_{cL.city='Bloomington'}(P \bowtie W \bowtie cL))\cap$$
$$\pi_{p.pid,p.pname}(\sigma_{p1.city='Chicago'}(P \bowtie W \bowtie cL \bowtie K \bowtie P1)))$$

Step 1: Pushing down selection over joins

$$\pi_{pid,pname}(\pi_{p.pid,pname}(P \bowtie W \bowtie \sigma_{cL.city='Bloomington'}(cL))\cap$$
$$\pi_{p.pid,p.pname}(P \bowtie W \bowtie cL \bowtie K\sigma_{p1.city='Chicago'}(P1)))$$

Step 2: Pushing down projections over joins

$\pi_{pid,pname}(\pi_{p.pid,pname}(P) \bowtie \pi_{w.pid,w.cname}(W) \bowtie \pi_{cL.canme}(\sigma_{cL.city='Bloomington'}(cL))\cap$
$\pi_{p.pid,p.pname}(P) \bowtie \pi_{w.pid,w.cname}(W) \bowtie\bowtie_{cL.cname}\ (cL) \bowtie \pi_{k.pid1,k.pid2}(K) \bowtie \pi_{pid}(\sigma_{p1.city='Chicago'}(P$

Step 3: Using semijoin

$\pi_{pid,pname}(\pi_{p.pid,pname}(P) \ltimes \pi_{w.pid,w.cname}(W) \ltimes \pi_{cL.cname}(\sigma_{cL.city='Bloomington'}(cL))\cap$
$\pi_{p.pid,p.pname}(P) \ltimes \pi_{w.pid,w.cname}(W) \ltimes (cL) \ltimes \pi_{k.pid1,k.pid2}(K) \ltimes \pi_{pid}(\sigma_{p1.city='Chicago'}(P1)))$

9.

$\pi_{c.cname,c.headquarter}(((C) \bowtie_{w.cname=c.cname}\ (W)) - Q$

$where,$
$Q = Q1 \cup Q2$
$Q1 = \pi_{q1.cname,q1.headquarter}(\pi_{w.pid,c.cname,c.headquarter}((\sigma_{w.salary \leq 70000}(W)) \bowtie (C) -$
$\pi_{w.pid,c.cname,c.headquarter}((\sigma_{w.salary \leq 70000}(W)) \bowtie C \bowtie_{w.pid=ps.pid}\ (\sigma_{pS.skill='Programming'}(pS)))$
$Q2 = \pi_{q1.cname,q1.headquarter}(\pi_{w.pid,c.cname,c.headquarter}((\sigma_{w.salary \leq 70000}(W)) \bowtie (C) -$
$\pi_{w.pid,c.cname,c.headquarter}((\sigma_{w.salary \leq 70000}(W)) \bowtie C \bowtie_{w.pid=ps.pid}\ (\sigma_{pS.skill='AI'}(pS)))$

Step 1: Pushing out projection

$\pi_{c.cname,c.headquarter}(((C) \bowtie_{w.cname=c.cname}\ (W)) - Q$

$where,$
$Q = Q1 \cup Q2$
$Q1 = \neg\pi_{w.pid,c.cname,c.headquarter}((\sigma_{w.salary \leq 70000}(W)) \bowtie C \bowtie_{w.pid=ps.pid}\ (\sigma_{pS.skill='Programming'}(pS)$
$Q2 = \neg\pi_{w.pid,c.cname,c.headquarter}((\sigma_{w.salary \leq 70000}(W)) \bowtie C \bowtie_{w.pid=ps.pid}\ (\sigma_{pS.skill='AI'}(pS)))$

Step 2:

$\pi_{c.cname,c.headquarter}(((C) \bowtie_{w.cname=c.cname}\ (W)) - Q$

$where,$
$Q = Q1 \cup Q2$
$Q1 = \neg\pi_{w.pid,c.cname,c.headquarter}(C \ltimes (\sigma_{w.salary \leq 70000})W \ltimes (\sigma_{w.pid=ps.pid})pS(\sigma_{pS.skill='Programming'})$
$Q2 = \neg\pi_{w.pid,c.cname,c.headquarter}(C \ltimes (\sigma_{w.salary \leq 70000})W \ltimes (\sigma_{w.pid=ps.pid})pS(\sigma_{pS.skill='AI'}))$

Step 3:

$\pi_{c.cname,c.headquarter}(((C) \bowtie_{w.cname=c.cname}\ (W)) - Q$

$where,$
$Q = \neg\pi_{w.pid,c.cname,c.headquarter}(C \ltimes (\sigma_{w.salary \leq 70000})W \ltimes (\sigma_{w.pid=ps.pid})pS(\sigma_{pS.skill='Programming'}))$
$\cup\neg\pi_{w.pid,c.cname,c.headquarter}(C \ltimes (\sigma_{w.salary \leq 70000})W \ltimes (\sigma_{w.pid=ps.pid})pS(\sigma_{pS.skill='AI'}))$

10.

$$\pi_{p.pid,T}((P) \bowtie_{hm1.mid=p.pid} (hM1) \bowtie_{hm2.mid=p.pid} (\sigma_{hm1.eid<>hm2.eid}(hM2)\cup$$
$$(\pi_{q.pid,F}(\pi_{p.pid}(P) - \pi_{p.pid}(P) \bowtie_{hm1.mid=p.pid} (hM1) \bowtie_{hm2.mid=p.pid} (\sigma_{hm1.eid<>hm2.eid}(hM2))))q)$$

Step 1:

$$\pi_{p.pid,T}((P) \bowtie_{hm1.mid=p.pid} (hM1) \bowtie_{hm2.mid=p.pid} (\sigma_{hm1.eid<>hm2.eid}(hM2)\cup$$
$$(\pi_{q.pid,F}(\neg\pi_{p.pid}(P) \bowtie_{hm1.mid=p.pid} (hM1) \bowtie_{hm2.mid=p.pid} (\sigma_{hm1.eid<>hm2.eid}(hM2))))q)$$