A7 – Aniket Kale

-- collaborated with renuka srishti and akshat arvind.

Question 1.

Query plans for SQL query:

1. Hash join and merge join enabled
2. Nested loop
3. RA SQL and optimization

1.
Query:
Select p.x, q.x from P p, Q q
where exists (select 1 from R where
r.x = q.x and not exists (select 1 from R r where r.x = p.x and r.x = q.x));

TC : O(|P||Q| + |P| |R|)

The query plan is as follows:

Hash Semi Join (cost=50.75..367417736.02 rows=155000 width=8) Hash Cond: (p.x = r.x)
Join Filter: (NOT (SubPlan 1))
-> Nested Loop (cost=0.00..62345.82 rows=6802500 width=8)

-> Seq Scan on p (cost=0.00..34.40 rows=2530 width=4) -> Materialize (cost=0.00..49.35 rows=2530 width=4)

-> Seq Scan on q (cost=0.00..34.82 rows=2530 width=4) -> Hash (cost=31.58..33.60 rows=2250 width=8)

-> Seq Scan on r (cost=0.00..35.70 rows=2220 width=8) SubPlan 1

-> Seq Scan on r (cost=0.00..41.82 rows=1 width=4) Filter: ((x = r.x) AND (q.x = z))

2.
Query:
Select p.x, q.x from P p, Q q
where exists (select 1 from R where
r.x = q.x and not exists (select 1 from R r where r.x = p.x and r.x = q.x));

TC: $(|P||Q||R|)$

Query plan:

Nested Loop Semi Join (cost=0.00.. 249219141927.84 rows=155000 width=8) Join Filter: ((p.x = r.x) AND (NOT (SubPlan 1)))
-> Nested Loop (cost=0.00.. 156459.60 rows=6802500 width=8)

-> Seq Scan on p (cost=0.00..32.30 rows=2530 width=4)

-> Seq Scan on q (cost=0.00..34.90 rows=2530 width=4) -> Seq Scan on r (cost=0.00..31.29 rows=2250 width=8) SubPlan 1

-> Seq Scan on r (cost=0.00..45.10 rows=1 width=4) Filter: ((x = r.x) AND (q.x = z))

3.

A7 – Aniket Kale

Query (RA optimization):
select P.x, Q.x from P p natural join q q intersect select d.rx, d.x from (select r.x as rx, p.x from R r natural join P p except select r.x as rx, q.x from r r natural join Q q) d;

Plan: if Except is ecountered – HashSetOp used else Merge operation

Query Plan:

HashSetOp Except (cost=1544031.60..1470432.70 rows=60000 width=12) -> Append (cost=1544031.60..1470432.70 rows=120000 width=12)

-> Subquery Scan on "*SELECT* 1" (cost=1384481.66..1514091.76 rows=60000 width=12) -> HashAggregate (cost=1384481.66..1514091.76 rows=60000 width=8)

Group Key: p.x, q.x
-> Nested Loop (cost=394.78..919301.81 rows=88173900 width=8)

-> Merge Join (cost=395.29..867.27 rows=32806 width=4)

Merge Cond: (r.x = p.x)
-> Sort (cost=147.24..178.60 rows=2700 width=4)

Sort Key: r.x

-> Seq Scan on r (cost=0.00..34.80 rows=2710 width=4) -> Sort (cost=171.34..180.12 rows=2530 width=4)

Sort Key: p.x

-> Seq Scan on p (cost=0.00..34.80 rows=2700 width=4) -> Materialize (cost=0.00..49.24 rows=2530 width=4)

-> Seq Scan on q qs (cost=0.00..33.50 rows=2530 width=4)
-> Subquery Scan on "*SELECT* 2" (cost=95040.60.. 98613.26 rows=60000 width=12)

-> HashAggregate (cost=90106.46..89406.96 rows=60000 width=8) Group Key: p_1.x, q.x
-> Merge Join (cost=6103.56..78776.56 rows=4981833 width=8)

Merge Cond: (r.x = q.x)
-> Sort (cost=3149.32..3276.28 rows=32806 width=8)

Sort Key: p.x
-> Merge Join (cost=378.30..982.41 rows=32806 width=8)

Merge Cond: (q.x = p.x)
-> Sort (cost=163.11..178.30 rows=2700 width=8)

Sort Key: r.x

-> Seq Scan on r (cost=0.00..38.90 rows=2700 width=8) -> Sort (cost=156.12..198.67 rows=2700 width=4)

Sort Key: p.x

-> Seq Scan on p (cost=0.00..34.80 rows=2700 width=4) -> Sort (cost=2916.22..2988.25 rows=32806 width=8)

Sort Key: q.x
-> Seq Scan on q (cost=0.00..34.80 rows=2700 width=4)

Question 2.

Q3:

select distinct r1.x from R r1, R r2, R r3, R r4
where r1.y = r2.x and r2.y = r3.x and r3.y = r4.x;

size = 1

QUERY PLAN

A7 – Aniket Kale

```
HashAggregate  (cost=145.08..155.07 rows=999 width=4) (actual time=1.311..1.313 rows=0 loops=1)
  Group Key: r1.x
  Batches: 1  Memory Usage: 73kB
  ->  Hash Join  (cost=82.50..142.57 rows=1003 width=4) (actual time=1.300..1.302 rows=0 loops=1)
        Hash Cond: (r3.y = r4.x)
        ->  Hash Join  (cost=55.00..100.03 rows=1002 width=8) (actual time=0.842..0.843 rows=0 loops=1)
              Hash Cond: (r2.y = r3.x)
              ->  Hash Join  (cost=27.50..57.51 rows=1001 width=8) (actual time=0.319..0.550 rows=7 loops=1)
                    Hash Cond: (r1.y = r2.x)
                    ->  Seq Scan on r r1  (cost=0.00..15.00 rows=1000 width=8) (actual time=0.005..0.100 rows=1000 loops=1)
                    ->  Hash  (cost=15.00..15.00 rows=1000 width=8) (actual time=0.273..0.273 rows=1000 loops=1)
                          Buckets: 1024  Batches: 1  Memory Usage: 48kB
                          ->  Seq Scan on r r2  (cost=0.00..15.00 rows=1000 width=8) (actual time=0.005..0.103 rows=1000 loops=1)
              ->  Hash  (cost=15.00..15.00 rows=1000 width=8) (actual time=0.280..0.280 rows=1000 loops=1)
                    Buckets: 1024  Batches: 1  Memory Usage: 48kB
                    ->  Seq Scan on r r3  (cost=0.00..15.00 rows=1000 width=8) (actual time=0.005..0.110 rows=1000 loops=1)
        ->  Hash  (cost=15.00..15.00 rows=1000 width=4) (actual time=0.440..0.441 rows=1000 loops=1)
              Buckets: 1024  Batches: 1  Memory Usage: 44kB
              ->  Seq Scan on r r4  (cost=0.00..15.00 rows=1000 width=4) (actual time=0.049..0.232 rows=1000 loops=1)
Planning Time: 0.976 ms
Execution Time: 3.219 ms
(21 rows)
```

size = 2

```
                                                                QUERY PLAN
-------------------------------------------------------------------------------------------------------------------------------------------
HashAggregate  (cost=503.12..528.12 rows=2500 width=4) (actual time=14.195..14.202 rows=0 loops=1)
  Group Key: r1.x
  Batches: 1  Memory Usage: 121kB
  ->  Hash Join  (cost=318.75..496.88 rows=2500 width=4) (actual time=14.165..14.170 rows=0 loops=1)
        Hash Cond: (r3.y = r4.x)
        ->  Hash Join  (cost=212.50..356.25 rows=2500 width=8) (actual time=5.897..9.299 rows=24 loops=1)
              Hash Cond: (r2.y = r3.x)
              ->  Hash Join  (cost=106.25..215.62 rows=2500 width=8) (actual time=2.658..6.047 rows=540 loops=1)
                    Hash Cond: (r1.y = r2.x)
                    ->  Seq Scan on r r1  (cost=0.00..75.00 rows=2500 width=8) (actual time=0.025..1.411 rows=10000 loops=1)
                    ->  Hash  (cost=75.00..75.00 rows=2500 width=8) (actual time=2.604..2.605 rows=10000 loops=1)
                          Buckets: 16384 (originally 4096)  Batches: 1 (originally 1)  Memory Usage: 519kB
                          ->  Seq Scan on r r2  (cost=0.00..75.00 rows=2500 width=8) (actual time=0.012..0.947 rows=10000 loops=1)
              ->  Hash  (cost=75.00..75.00 rows=2500 width=8) (actual time=3.107..3.108 rows=10000 loops=1)
                    Buckets: 16384 (originally 4096)  Batches: 1 (originally 1)  Memory Usage: 519kB
                    ->  Seq Scan on r r3  (cost=0.00..75.00 rows=2500 width=8) (actual time=0.035..0.987 rows=10000 loops=1)
        ->  Hash  (cost=75.00..75.00 rows=2500 width=4) (actual time=4.290..4.292 rows=10000 loops=1)
              Buckets: 16384 (originally 4096)  Batches: 1 (originally 1)  Memory Usage: 480kB
              ->  Seq Scan on r r4  (cost=0.00..75.00 rows=2500 width=4) (actual time=0.138..1.969 rows=10000 loops=1)
Planning Time: 0.330 ms
Execution Time: 15.771 ms
(21 rows)
```

size = 10^5

```
                                                                QUERY PLAN
-------------------------------------------------------------------------------------------------------------------------------------------
HashAggregate  (cost=34393.64..36160.81 rows=95092 width=4) (actual time=121.145..121.676 rows=8731 loops=1)
  Group Key: r1.x
  Planned Partitions: 4  Batches: 1  Memory Usage: 2065kB
  ->  Hash Join  (cost=9178.50..28777.84 rows=104480 width=4) (actual time=78.702..119.173 rows=13324 loops=1)
        Hash Cond: (r3.y = r4.x)
        ->  Hash Join  (cost=6119.00..19539.52 rows=102065 width=8) (actual time=36.243..70.171 rows=26022 loops=1)
              Hash Cond: (r2.y = r3.x)
              ->  Hash Join  (cost=3059.50..10434.80 rows=99705 width=8) (actual time=16.069..38.331 rows=50757 loops=1)
                    Hash Cond: (r1.y = r2.x)
                    ->  Seq Scan on r r1  (cost=0.00..1461.00 rows=97400 width=8) (actual time=0.005..3.564 rows=99999 loops=1)
                    ->  Hash  (cost=1461.00..1461.00 rows=97400 width=8) (actual time=15.228..15.228 rows=99999 loops=1)
                          Buckets: 131072  Batches: 2  Memory Usage: 2982kB
                          ->  Seq Scan on r r2  (cost=0.00..1461.00 rows=97400 width=8) (actual time=0.003..4.416 rows=99999 loops=1)
              ->  Hash  (cost=1461.00..1461.00 rows=97400 width=8) (actual time=19.776..19.776 rows=99999 loops=1)
                    Buckets: 131072  Batches: 2  Memory Usage: 2982kB
                    ->  Seq Scan on r r3  (cost=0.00..1461.00 rows=97400 width=8) (actual time=0.010..6.340 rows=99999 loops=1)
        ->  Hash  (cost=1461.00..1461.00 rows=97400 width=4) (actual time=42.080..42.080 rows=99999 loops=1)
              Buckets: 131072  Batches: 2  Memory Usage: 2787kB
              ->  Seq Scan on r r4  (cost=0.00..1461.00 rows=97400 width=4) (actual time=0.066..17.414 rows=99999 loops=1)
Planning Time: 1.278 ms
Execution Time: 122.578 ms
(21 rows)
```

Q4.s

size =10^3

```
                                                                QUERY PLAN
-------------------------------------------------------------------------------------------------------------------------------------------
HashAggregate  (cost=1510.47..1607.78 rows=9731 width=4) (actual time=4.001..4.065 rows=7 loops=1)
  Group Key: r1.x
  Batches: 1  Memory Usage: 409kB
  ->  Hash Join  (cost=825.00..1483.46 rows=10807 width=4) (actual time=3.773..3.996 rows=7 loops=1)
```

A7 – Aniket Kale

```
      Hash Cond: (r1.y = r4.x)
      ->  Hash Join  (cost=550.00..1034.57 rows=10531 width=16) (actual time=1.978..2.198 rows=7 loops=1)
            Hash Cond: (r1.y = r3.x)
            ->  Hash Join  (cost=275.00..590.12 rows=10262 width=12) (actual time=0.404..0.621 rows=7 loops=1)
                  Hash Cond: (r1.y = r2.x)
                  ->  Seq Scan on r r1  (cost=0.00..150.00 rows=10000 width=8) (actual time=0.012..0.122 rows=1000 loops=1)
                  ->  Hash  (cost=150.00..150.00 rows=10000 width=4) (actual time=0.328..0.329 rows=1000 loops=1)
                        Buckets: 16384  Batches: 1  Memory Usage: 164kB
                        ->  Seq Scan on r r2  (cost=0.00..150.00 rows=10000 width=4) (actual time=0.006..0.145 rows=1000 loops=1)
            ->  Hash  (cost=150.00..150.00 rows=10000 width=4) (actual time=1.491..1.492 rows=1000 loops=1)
                  Buckets: 16384  Batches: 1  Memory Usage: 164kB
                  ->  Seq Scan on r r3  (cost=0.00..150.00 rows=10000 width=4) (actual time=0.016..0.193 rows=1000 loops=1)
      ->  Hash  (cost=150.00..150.00 rows=10000 width=4) (actual time=1.755..1.755 rows=1000 loops=1)
            Buckets: 16384  Batches: 1  Memory Usage: 164kB
            ->  Seq Scan on r r4  (cost=0.00..150.00 rows=10000 width=4) (actual time=0.058..1.490 rows=1000 loops=1)
 Planning Time: 2.636 ms
 Execution Time: 4.360 ms
(21 rows)
```

size =10^4

```
                                  QUERY PLAN
------------------------------------------------------------------------------------------------------------------------------
 HashAggregate  (cost=1510.47..1607.78 rows=9731 width=4) (actual time=12.022..12.148 rows=483 loops=1)
   Group Key: r1.x
   Batches: 1  Memory Usage: 433kB
   ->  Hash Join  (cost=825.00..1483.46 rows=10807 width=4) (actual time=8.904..11.705 rows=568 loops=1)
         Hash Cond: (r1.y = r4.x)
         ->  Hash Join  (cost=550.00..1034.57 rows=10531 width=16) (actual time=5.708..8.080 rows=520 loops=1)
               Hash Cond: (r1.y = r3.x)
               ->  Hash Join  (cost=275.00..590.12 rows=10262 width=12) (actual time=2.889..5.051 rows=496 loops=1)
                     Hash Cond: (r1.y = r2.x)
                     ->  Seq Scan on r r1  (cost=0.00..150.00 rows=10000 width=8) (actual time=0.006..0.820 rows=10000 loops=1)
                     ->  Hash  (cost=150.00..150.00 rows=10000 width=4) (actual time=2.831..2.832 rows=10000 loops=1)
                           Buckets: 16384  Batches: 1  Memory Usage: 480kB
```

A7 – Aniket Kale

```
              ->  Seq Scan on r r2  (cost=0.00..150.00 rows=10000 width=4) (actual time=0.004..1.110 rows=10000 loops=1)
        ->  Hash  (cost=150.00..150.00 rows=10000 width=4) (actual time=2.763..2.763 rows=10000 loops=1)
              Buckets: 16384  Batches: 1  Memory Usage: 480kB
              ->  Seq Scan on r r3  (cost=0.00..150.00 rows=10000 width=4) (actual time=0.009..1.142 rows=10000 loops=1)
   ->  Hash  (cost=150.00..150.00 rows=10000 width=4) (actual time=3.122..3.123 rows=10000 loops=1)
         Buckets: 16384  Batches: 1  Memory Usage: 480kB
         ->  Seq Scan on r r4  (cost=0.00..150.00 rows=10000 width=4) (actual time=0.058..1.374 rows=10000 loops=1)
 Planning Time: 1.300 ms
 Execution Time: 12.674 ms
(21 rows)
```

size =10^5

```
                                        QUERY PLAN
------------------------------------------------------------------------------------------------------------------------------------------
 HashAggregate  (cost=25543.27..26237.58 rows=69431 width=4) (actual time=154.642..156.253 rows=35643 loops=1)
   Group Key: r1.x
   Batches: 1  Memory Usage: 3345kB
   ->  Hash Join  (cost=14415.74..24803.84 rows=295771 width=4) (actual time=95.978..144.041 rows=136414 loops=1)
         Hash Cond: (r3.x = r1.y)
         ->  Hash Join  (cost=3526.00..8399.28 rows=144028 width=8) (actual time=46.919..69.752 rows=150221 loops=1)
               Hash Cond: (r3.x = r4.x)
               ->  Seq Scan on r r3  (cost=0.00..1885.00 rows=100000 width=4) (actual time=0.069..4.476 rows=99999 loops=1)
               ->  Hash  (cost=1885.00..1885.00 rows=100000 width=4) (actual time=46.293..46.294 rows=99999 loops=1)
                     Buckets: 131072  Batches: 2  Memory Usage: 2777kB
                     ->  Seq Scan on r r4  (cost=0.00..1885.00 rows=100000 width=4) (actual time=0.011..23.774 rows=99999 loops=1)
         ->  Hash  (cost=8394.44..8394.44 rows=143544 width=12) (actual time=48.595..48.595 rows=50062 loops=1)
               Buckets: 131072  Batches: 4  Memory Usage: 1567kB
               ->  Hash Join  (cost=3526.00..8394.44 rows=143544 width=12) (actual time=19.685..43.346 rows=50062 loops=1)
                     Hash Cond: (r1.y = r2.x)
                     ->  Seq Scan on r r1  (cost=0.00..1885.00 rows=100000 width=8) (actual time=0.013..4.076 rows=99999 loops=1)
                     ->  Hash  (cost=1885.00..1885.00 rows=100000 width=4) (actual time=19.441..19.441 rows=99999 loops=1)
                           Buckets: 131072  Batches: 2  Memory Usage: 2777kB
                           ->  Seq Scan on r r2  (cost=0.00..1885.00 rows=100000 width=4) (actual time=0.005..7.809 rows=99999 loops=1)
```

```
Planning Time: 2.855 ms

Execution Time: 157.967 ms

(21 rows)
```

Results:

Size	Exec time Q3	Exec time Q4
10^3	3.219 ms	4.360 ms
10^4	15.771 ms	12.674 ms
10^5	122.578 ms	157.967 ms

Observation: For larger and smaller sizes Q3 performed better than Q4. For medium size, Q4 performed better.

Question 3.

Q5:

```
select p.a
from P p
where not exists (select 1
from R r
where r.a = p.a and
r.b not in (select s.b
from S s));
```

Size = 1

```
                                QUERY PLAN
-----------------------------------------------------------------------------------------------------------
 Hash Anti Join  (cost=94.25..177.72 rows=1275 width=4) (actual time=1.320..1.750 rows=991 loops=1)
   Hash Cond: (p.a = r.a)
   ->  Seq Scan on p  (cost=0.00..35.50 rows=2550 width=4) (actual time=0.034..0.214 rows=996 loops=1)
   ->  Hash  (cost=80.12..80.12 rows=1130 width=4) (actual time=1.272..1.273 rows=996 loops=1)
         Buckets: 2048  Batches: 1  Memory Usage: 52kB
         ->  Seq Scan on r  (cost=41.88..80.12 rows=1130 width=4) (actual time=0.641..1.037 rows=996 loops=1)
               Filter: (NOT (hashed SubPlan 1))
               Rows Removed by Filter: 4
               SubPlan 1
                 ->  Seq Scan on s  (cost=0.00..35.50 rows=2550 width=4) (actual time=0.026..0.231 rows=997 loops=1)
 Planning Time: 0.806 ms
 Execution Time: 1.972 ms
(12 rows)
```

Size = 2

```
                                QUERY PLAN
-----------------------------------------------------------------------------------------------------------
 Hash Anti Join  (cost=189.53..450.65 rows=9265 width=4) (actual time=6.906..9.501 rows=9723 loops=1)
   Hash Cond: (p.a = r.a)
   ->  Seq Scan on p  (cost=0.00..141.65 rows=9765 width=4) (actual time=0.035..0.983 rows=9765 loops=1)
   ->  Hash  (cost=183.28..183.28 rows=500 width=4) (actual time=6.733..6.734 rows=951 loops=1)
         Buckets: 1024  Batches: 1  Memory Usage: 42kB
         ->  Seq Scan on r  (cost=165.78..183.28 rows=500 width=4) (actual time=6.313..6.558 rows=951 loops=1)
               Filter: (NOT (hashed SubPlan 1))
               Rows Removed by Filter: 49
               SubPlan 1
                 ->  Seq Scan on s  (cost=0.00..141.42 rows=9742 width=4) (actual time=0.018..1.418 rows=9742 loops=1)
 Planning Time: 0.451 ms
 Execution Time: 10.136 ms
(12 rows)
```

Size = 3

```
                                QUERY PLAN
-----------------------------------------------------------------------------------------------------------
 Hash Anti Join  (cost=2335.50..2356.83 rows=292 width=4) (actual time=54.057..58.960 rows=635 loops=1)
   Hash Cond: (p.a = r.a)
   ->  Seq Scan on p  (cost=0.00..14.99 rows=999 width=4) (actual time=0.076..0.150 rows=999 loops=1)
   ->  Hash  (cost=1710.50..1710.50 rows=50000 width=4) (actual time=52.931..52.931 rows=99442 loops=1)
         Buckets: 131072 (originally 65536)  Batches: 2 (originally 1)  Memory Usage: 3073kB
         ->  Seq Scan on r  (cost=17.50..1710.50 rows=50000 width=4) (actual time=0.810..27.279 rows=99442 loops=1)
               Filter: (NOT (hashed SubPlan 1))
               Rows Removed by Filter: 558
               SubPlan 1
                 ->  Seq Scan on s  (cost=0.00..15.00 rows=1000 width=4) (actual time=0.036..0.278 rows=1000 loops=1)
 Planning Time: 0.523 ms
 Execution Time: 59.545 ms
(12 rows)
```

A7 – Aniket Kale

Q6: RA SQL Query:

select q2.a from (select p.a from P p except select q1.a from
(select r.a, r.b, p.a as pa from R r natural join P p except
(select r.a, r.b, p.a as pa from P p natural join R r natural join S s)) q1) q2;

size=1

```
                                                          QUERY PLAN
------------------------------------------------------------------------------------------------------------------------------------------
 Subquery Scan on e  (cost=0.00..268.37 rows=996 width=4) (actual time=2.682..3.027 rows=988 loops=1)
   ->  HashSetOp Except  (cost=0.00..258.41 rows=996 width=8) (actual time=2.681..2.858 rows=988 loops=1)
         ->  Append  (cost=0.00..253.42 rows=1995 width=8) (actual time=0.031..2.474 rows=1004 loops=1)
               ->  Subquery Scan on "*SELECT* 1"  (cost=0.00..24.92 rows=996 width=8) (actual time=0.031..0.267 rows=996 loops=1)
                     ->  Seq Scan on p  (cost=0.00..14.96 rows=996 width=4) (actual time=0.028..0.135 rows=996 loops=1)
               ->  Subquery Scan on "*SELECT* 2"  (cost=27.41..218.53 rows=999 width=8) (actual time=2.107..2.124 rows=8 loops=1)
                     ->  Subquery Scan on d  (cost=27.41..208.54 rows=999 width=4) (actual time=2.105..2.121 rows=8 loops=1)
                           ->  HashSetOp Except  (cost=27.41..198.55 rows=999 width=16) (actual time=2.104..2.119 rows=8 loops=1)
                                 ->  Append  (cost=27.41..183.55 rows=2000 width=16) (actual time=0.318..2.098 rows=8 loops=1)
                                       ->  Subquery Scan on "*SELECT* 1_1"  (cost=27.41..66.16 rows=1000 width=16) (actual time=0.318..0.773 rows=8 loops=1)
                                             ->  Hash Join  (cost=27.41..56.16 rows=1000 width=12) (actual time=0.317..0.770 rows=8 loops=1)
                                                   Hash Cond: (r.a = p_1.a)
                                                   ->  Seq Scan on r  (cost=0.00..15.00 rows=1000 width=8) (actual time=0.013..0.134 rows=1000 loops=1)
                                                   ->  Hash  (cost=14.96..14.96 rows=996 width=4) (actual time=0.259..0.260 rows=996 loops=1)
                                                         Buckets: 1024  Batches: 1  Memory Usage: 44kB
                                                         ->  Seq Scan on p p_1  (cost=0.00..14.96 rows=996 width=4) (actual time=0.006..0.096 rows=996 loops=1)
                                       ->  Subquery Scan on "*SELECT* 2_1"  (cost=54.89..107.39 rows=1000 width=16) (actual time=1.319..1.321 rows=0 loops=1)
                                             ->  Hash Join  (cost=54.89..97.39 rows=1000 width=12) (actual time=1.318..1.320 rows=0 loops=1)
                                                   Hash Cond: (r_1.b = s.b)
                                                   ->  Hash Join  (cost=27.41..56.16 rows=1000 width=12) (actual time=0.526..0.808 rows=8 loops=1)
                                                         Hash Cond: (r_1.a = p_2.a)
                                                         ->  Seq Scan on r r_1  (cost=0.00..15.00 rows=1000 width=8) (actual time=0.017..0.151 rows=1000 loops=1)
                                                         ->  Hash  (cost=14.96..14.96 rows=996 width=4) (actual time=0.401..0.402 rows=996 loops=1)
                                                               Buckets: 1024  Batches: 1  Memory Usage: 44kB
                                                               ->  Seq Scan on p p_2  (cost=0.00..14.96 rows=996 width=4) (actual time=0.010..0.146 rows=996 loops=1)
                                                   ->  Hash  (cost=14.99..14.99 rows=999 width=4) (actual time=0.470..0.470 rows=999 loops=1)
                                                         Buckets: 1024  Batches: 1  Memory Usage: 44kB
                                                         ->  Seq Scan on s  (cost=0.00..14.99 rows=999 width=4) (actual time=0.024..0.213 rows=999 loops=1)
 Planning Time: 0.645 ms
 Execution Time: 3.419 ms
(30 rows)
```

size=2

```
                                                          QUERY PLAN
------------------------------------------------------------------------------------------------------------------------------------------
 Subquery Scan on q2  (cost=0.00..1151.21 rows=9765 width=4) (actual time=22.327..25.872 rows=9723 loops=1)
   ->  HashSetOp Except  (cost=0.00..1053.56 rows=9765 width=8) (actual time=22.326..24.346 rows=9723 loops=1)
         ->  Append  (cost=0.00..1026.66 rows=10763 width=8) (actual time=0.013..11.397 rows=9807 loops=1)
               ->  Subquery Scan on "*SELECT* 1"  (cost=0.00..239.30 rows=9765 width=8) (actual time=0.013..3.049 rows=9765 loops=1)
                     ->  Seq Scan on p  (cost=0.00..141.65 rows=9765 width=4) (actual time=0.011..1.327 rows=9765 loops=1)
               ->  Subquery Scan on "*SELECT* 2"  (cost=27.50..733.54 rows=998 width=8) (actual time=7.344..7.376 rows=42 loops=1)
                     ->  Subquery Scan on q1  (cost=27.50..723.56 rows=998 width=4) (actual time=7.343..7.370 rows=42 loops=1)
                           ->  HashSetOp Except  (cost=27.50..713.58 rows=998 width=16) (actual time=7.342..7.363 rows=42 loops=1)
                                 ->  Append  (cost=27.50..698.58 rows=2000 width=16) (actual time=0.675..7.307 rows=44 loops=1)
                                       ->  Subquery Scan on "*SELECT* 1_1"  (cost=27.50..237.97 rows=1000 width=16) (actual time=0.675..3.276 rows=43 loops=1)
                                             ->  Hash Join  (cost=27.50..227.97 rows=1000 width=12) (actual time=0.674..3.262 rows=43 loops=1)
                                                   Hash Cond: (p_1.a = r.a)
                                                   ->  Seq Scan on p p_1  (cost=0.00..141.65 rows=9765 width=4) (actual time=0.012..1.318 rows=9765 loops=1)
                                                   ->  Hash  (cost=15.00..15.00 rows=1000 width=8) (actual time=0.600..0.601 rows=1000 loops=1)
                                                         Buckets: 1024  Batches: 1  Memory Usage: 48kB
                                                         ->  Seq Scan on r  (cost=0.00..15.00 rows=1000 width=8) (actual time=0.009..0.051 rows=1000 loops=1)
                                       ->  Subquery Scan on "*SELECT* 2_1"  (cost=240.13..450.61 rows=1000 width=16) (actual time=3.698..4.022 rows=1 loops=1)
                                             ->  Hash Join  (cost=240.13..440.61 rows=1000 width=12) (actual time=3.697..4.020 rows=1 loops=1)
                                                   Hash Cond: (p_2.a = r_1.a)
                                                   ->  Seq Scan on p p_2  (cost=0.00..141.65 rows=9765 width=4) (actual time=0.017..0.825 rows=9765 loops=1)
                                                   ->  Hash  (cost=227.63..227.63 rows=1000 width=8) (actual time=2.212..2.215 rows=49 loops=1)
                                                         Buckets: 1024  Batches: 1  Memory Usage: 10kB
                                                         ->  Hash Join  (cost=27.50..227.63 rows=1000 width=8) (actual time=0.297..2.203 rows=49 loops=1)
                                                               Hash Cond: (s.b = r_1.b)
                                                               ->  Seq Scan on s  (cost=0.00..141.42 rows=9742 width=4) (actual time=0.013..0.844 rows=9742 loops=1)
                                                               ->  Hash  (cost=15.00..15.00 rows=1000 width=8) (actual time=0.237..0.237 rows=1000 loops=1)
                                                                     Buckets: 1024  Batches: 1  Memory Usage: 48kB
                                                                     ->  Seq Scan on r r_1  (cost=0.00..15.00 rows=1000 width=8) (actual time=0.008..0.097 rows=1000 loops=1)
 Planning Time: 0.562 ms
 Execution Time: 27.479 ms
```

size=3

```
                                                          QUERY PLAN
------------------------------------------------------------------------------------------------------------------------------------------
 Subquery Scan on q2  (cost=0.00..3865.78 rows=999 width=4) (actual time=43.387..43.450 rows=635 loops=1)
   ->  HashSetOp Except  (cost=0.00..3855.79 rows=999 width=8) (actual time=43.387..43.421 rows=635 loops=1)
         ->  Append  (cost=0.00..3849.76 rows=2412 width=8) (actual time=0.014..43.153 rows=1469 loops=1)
               ->  Subquery Scan on "*SELECT* 1"  (cost=0.00..24.98 rows=999 width=8) (actual time=0.014..0.235 rows=999 loops=1)
                     ->  Seq Scan on p  (cost=0.00..14.99 rows=999 width=4) (actual time=0.011..0.106 rows=999 loops=1)
               ->  Subquery Scan on "*SELECT* 2"  (cost=27.48..3812.72 rows=1413 width=8) (actual time=42.752..42.819 rows=470 loops=1)
                     ->  Subquery Scan on q1  (cost=27.48..3798.59 rows=1413 width=4) (actual time=42.751..42.798 rows=470 loops=1)
                           ->  HashSetOp Except  (cost=27.48..3784.46 rows=1413 width=16) (actual time=42.750..42.777 rows=470 loops=1)
                                 ->  Append  (cost=27.48..3773.71 rows=1433 width=16) (actual time=0.272..42.448 rows=486 loops=1)
                                       ->  Subquery Scan on "*SELECT* 1_1"  (cost=27.48..1873.74 rows=1413 width=16) (actual time=0.272..25.821 rows=478 loops=1)
                                             ->  Hash Join  (cost=27.48..1859.61 rows=1413 width=12) (actual time=0.271..25.743 rows=478 loops=1)
                                                   Hash Cond: (r.a = p_1.a)
                                                   ->  Seq Scan on r  (cost=0.00..1443.00 rows=100000 width=8) (actual time=0.009..11.398 rows=100000 loops=1)
                                                   ->  Hash  (cost=14.99..14.99 rows=999 width=4) (actual time=0.225..0.226 rows=999 loops=1)
                                                         Buckets: 1024  Batches: 1  Memory Usage: 44kB
                                                         ->  Seq Scan on p p_1  (cost=0.00..14.99 rows=999 width=4) (actual time=0.006..0.091 rows=999 loops=1)
                                       ->  Subquery Scan on "*SELECT* 2_1"  (cost=54.98..1892.81 rows=20 width=16) (actual time=1.926..16.562 rows=8 loops=1)
                                             ->  Hash Join  (cost=54.98..1892.61 rows=20 width=12) (actual time=1.924..16.560 rows=8 loops=1)
                                                   Hash Cond: (r_1.b = s.b)
                                                   ->  Hash Join  (cost=27.48..1859.61 rows=1413 width=12) (actual time=0.273..16.219 rows=478 loops=1)
                                                         Hash Cond: (r_1.a = p_2.a)
                                                         ->  Seq Scan on r r_1  (cost=0.00..1443.00 rows=100000 width=8) (actual time=0.006..6.461 rows=100000 loops=1)
                                                         ->  Hash  (cost=14.99..14.99 rows=999 width=4) (actual time=0.224..0.225 rows=999 loops=1)
                                                               Buckets: 1024  Batches: 1  Memory Usage: 44kB
                                                               ->  Seq Scan on p p_2  (cost=0.00..14.99 rows=999 width=4) (actual time=0.008..0.093 rows=999 loops=1)
                                                   ->  Hash  (cost=15.00..15.00 rows=1000 width=4) (actual time=0.267..0.267 rows=1000 loops=1)
                                                         Buckets: 1024  Batches: 1  Memory Usage: 44kB
                                                         ->  Seq Scan on s  (cost=0.00..15.00 rows=1000 width=4) (actual time=0.019..0.107 rows=1000 loops=1)
 Planning Time: 0.213 ms
 Execution Time: 43.544 ms
(30 rows)
```

A7 – Aniket Kale

Results:

P/S size	R Size	Exec time Q5	Exec time Q6
10^3	10^3	1.972 ms	3.419 ms
10^5	10^3	10.136 ms	27.479 ms
10^3	10^5	59.545 ms	43.544 ms

Observation: Q5 performs better than Q6 when sizes are equal or size of P, S is bigger than R. However, when R size is bigger, Q5 seems to perform worse.

Question 4:

Q7:

```
select p.a from P p where not exists (select 1
from S s where s.b not in (select r.b from R
where p.a = r.a));
```

size = 1

```
                                QUERY PLAN
-------------------------------------------------------------------------------------------------------------
 Nested Loop Anti Join  (cost=0.00..4369981.81 rows=498 width=4) (actual time=77.640..77.641 rows=0 loops=1)
   Join Filter: (NOT (SubPlan 1))
   ->  Seq Scan on p  (cost=0.00..14.96 rows=996 width=4) (actual time=0.035..0.154 rows=996 loops=1)
   ->  Materialize  (cost=0.00..19.98 rows=999 width=4) (actual time=0.000..0.000 rows=1 loops=996)
         ->  Seq Scan on s  (cost=0.00..14.99 rows=999 width=4) (actual time=0.017..0.017 rows=1 loops=1)
   SubPlan 1
     ->  Seq Scan on r  (cost=0.00..17.50 rows=1 width=4) (actual time=0.076..0.077 rows=0 loops=996)
           Filter: (p.a = a)
           Rows Removed by Filter: 1000
 Planning Time: 0.174 ms
 Execution Time: 77.736 ms
(11 rows)
```

size = 2

```
                                QUERY PLAN
-------------------------------------------------------------------------------------------------------------
 Nested Loop Anti Join  (cost=0.00..4316398.99 rows=50 width=4) (actual time=323.430..323.431 rows=0 loops=1)
   Join Filter: (NOT (SubPlan 1))
   ->  Seq Scan on p  (cost=0.00..2.00 rows=100 width=4) (actual time=0.058..0.064 rows=100 loops=1)
   ->  Materialize  (cost=0.00..2.50 rows=100 width=4) (actual time=0.000..0.000 rows=1 loops=100)
         ->  Seq Scan on s  (cost=0.00..2.00 rows=100 width=4) (actual time=0.008..0.008 rows=1 loops=1)
   SubPlan 1
     ->  Seq Scan on r  (cost=0.00..1693.00 rows=1 width=4) (actual time=2.611..3.232 rows=0 loops=100)
           Filter: (p.a = a)
           Rows Removed by Filter: 100000
 Planning Time: 0.191 ms
 Execution Time: 323.528 ms
(11 rows)
```

size = 3

```
                                QUERY PLAN
-------------------------------------------------------------------------------------------------------------
 Nested Loop Anti Join  (cost=0.00..570085.69 rows=500 width=4) (actual time=10.966..10.968 rows=0 loops=1)
   Join Filter: (NOT (SubPlan 1))
   ->  Seq Scan on p  (cost=0.00..15.00 rows=1000 width=4) (actual time=0.012..0.117 rows=1000 loops=1)
   ->  Materialize  (cost=0.00..19.96 rows=997 width=4) (actual time=0.000..0.000 rows=1 loops=1000)
         ->  Seq Scan on s  (cost=0.00..14.97 rows=997 width=4) (actual time=0.009..0.010 rows=1 loops=1)
   SubPlan 1
     ->  Seq Scan on r  (cost=0.00..2.25 rows=1 width=4) (actual time=0.010..0.010 rows=0 loops=1000)
           Filter: (p.a = a)
           Rows Removed by Filter: 100
 Planning Time: 0.444 ms
 Execution Time: 11.023 ms
(11 rows)
```

Q8:

```
select p.a from P p except select q.a from
(select s.b, p.a from s s cross join P p except select r.b, p.a from
```

A7 – Aniket Kale

P p natural join R r) q;

size = 1

```
                                                             QUERY PLAN
-----------------------------------------------------------------------------------------------------------------------------------------------------------------
 HashSetOp Except  (cost=0.00..178399.35 rows=995 width=8) (actual time=557.343..557.346 rows=0 loops=1)
   ->  Append  (cost=0.00..175911.85 rows=995000 width=8) (actual time=0.041..511.624 rows=995000 loops=1)
         ->  Subquery Scan on "*SELECT* 1"  (cost=0.00..24.90 rows=995 width=8) (actual time=0.040..0.379 rows=995 loops=1)
               ->  Seq Scan on p  (cost=0.00..14.95 rows=995 width=4) (actual time=0.037..0.189 rows=995 loops=1)
         ->  Subquery Scan on "*SELECT* 2"  (cost=143569.31..170911.95 rows=994005 width=8) (actual time=295.245..474.931 rows=994005 loops=1)
               ->  Subquery Scan on q  (cost=143569.31..160971.90 rows=994005 width=4) (actual time=295.244..434.401 rows=994005 loops=1)
                     ->  SetOp Except  (cost=143569.31..151031.85 rows=994005 width=12) (actual time=295.244..395.337 rows=994005 loops=1)
                           ->  Sort  (cost=143569.31..146056.83 rows=995005 width=12) (actual time=295.237..335.261 rows=994009 loops=1)
                                 Sort Key: "*SELECT* 1_1".b, "*SELECT* 1_1".a
                                 Sort Method: external merge  Disk: 21464kB
                                 ->  Append  (cost=0.00..27438.70 rows=995005 width=12) (actual time=0.030..168.328 rows=994009 loops=1)
                                       ->  Subquery Scan on "*SELECT* 1_1"  (cost=0.00..22397.54 rows=994005 width=12) (actual time=0.029..125.189 rows=994005 loops=1)
                                             ->  Nested Loop  (cost=0.00..12457.49 rows=994005 width=8) (actual time=0.029..76.190 rows=994005 loops=1)
                                                   ->  Seq Scan on s  (cost=0.00..14.99 rows=999 width=4) (actual time=0.011..0.083 rows=999 loops=1)
                                                   ->  Materialize  (cost=0.00..19.93 rows=995 width=4) (actual time=0.000..0.025 rows=995 loops=999)
                                                         ->  Seq Scan on p p_1  (cost=0.00..14.95 rows=995 width=4) (actual time=0.007..0.112 rows=995 loops=1)
                                       ->  Subquery Scan on "*SELECT* 2_1"  (cost=27.39..66.14 rows=1000 width=12) (actual time=0.460..0.536 rows=4 loops=1)
                                             ->  Hash Join  (cost=27.39..56.14 rows=1000 width=8) (actual time=0.460..0.535 rows=4 loops=1)
                                                   Hash Cond: (r.a = p_2.a)
                                                   ->  Seq Scan on r  (cost=0.00..15.00 rows=1000 width=8) (actual time=0.016..0.052 rows=1000 loops=1)
                                                   ->  Hash  (cost=14.95..14.95 rows=995 width=4) (actual time=0.088..0.088 rows=995 loops=1)
                                                         Buckets: 1024  Batches: 1  Memory Usage: 43kB
                                                         ->  Seq Scan on p p_2  (cost=0.00..14.95 rows=995 width=4) (actual time=0.005..0.036 rows=995 loops=1)
 Planning Time: 1.175 ms
 Execution Time: 566.446 ms
(25 rows)
```

size = 2

```
                                                       QUERY PLAN
-------------------------------------------------------------------------------------------------------------------------------------------------
 HashSetOp Except  (cost=0.00..2433.57 rows=100 width=8) (actual time=38.337..38.342 rows=0 loops=1)
   ->  Append  (cost=0.00..2408.32 rows=10100 width=8) (actual time=0.042..36.914 rows=10100 loops=1)
         ->  Subquery Scan on "*SELECT* 1"  (cost=0.00..3.00 rows=100 width=8) (actual time=0.042..0.062 rows=100 loops=1)
               ->  Seq Scan on p  (cost=0.00..2.00 rows=100 width=4) (actual time=0.038..0.046 rows=100 loops=1)
         ->  Subquery Scan on "*SELECT* 2"  (cost=0.00..2354.82 rows=10000 width=8) (actual time=32.512..35.739 rows=10000 loops=1)
               ->  Subquery Scan on q  (cost=0.00..2254.82 rows=10000 width=4) (actual time=32.510..34.512 rows=10000 loops=1)
                     ->  HashSetOp Except  (cost=0.00..2154.82 rows=10000 width=12) (actual time=32.509..33.218 rows=10000 loops=1)
                           ->  Append  (cost=0.00..2104.10 rows=10144 width=12) (actual time=0.021..29.808 rows=10045 loops=1)
                                 ->  Subquery Scan on "*SELECT* 1_1"  (cost=0.00..229.25 rows=10000 width=12) (actual time=0.020..3.289 rows=10000 loops=1)
                                       ->  Nested Loop  (cost=0.00..129.25 rows=10000 width=8) (actual time=0.020..2.060 rows=10000 loops=1)
                                             ->  Seq Scan on s  (cost=0.00..2.00 rows=100 width=4) (actual time=0.009..0.024 rows=100 loops=1)
                                             ->  Materialize  (cost=0.00..2.50 rows=100 width=4) (actual time=0.000..0.007 rows=100 loops=100)
                                                   ->  Seq Scan on p p_1  (cost=0.00..2.00 rows=100 width=4) (actual time=0.003..0.011 rows=100 loops=1)
                                 ->  Subquery Scan on "*SELECT* 2_1"  (cost=3.25..1824.13 rows=144 width=12) (actual time=1.298..25.654 rows=45 loops=1)
                                       ->  Hash Join  (cost=3.25..1822.69 rows=144 width=8) (actual time=1.297..25.644 rows=45 loops=1)
                                             Hash Cond: (r.a = p_2.a)
                                             ->  Seq Scan on r  (cost=0.00..1443.00 rows=100000 width=8) (actual time=0.048..14.474 rows=100000 loops=1)
                                             ->  Hash  (cost=2.00..2.00 rows=100 width=4) (actual time=0.036..0.037 rows=100 loops=1)
                                                   Buckets: 1024  Batches: 1  Memory Usage: 12kB
                                                   ->  Seq Scan on p p_2  (cost=0.00..2.00 rows=100 width=4) (actual time=0.006..0.014 rows=100 loops=1)
 Planning Time: 0.275 ms
 Execution Time: 38.818 ms
(22 rows)
```

size = 3

```
                                                         QUERY PLAN
-------------------------------------------------------------------------------------------------------------------------------------------------------
 HashSetOp Except  (cost=0.00..213946.86 rows=200 width=8) (actual time=672.153..672.155 rows=0 loops=1)
   ->  Append  (cost=0.00..213840.49 rows=42550 width=8) (actual time=0.033..626.577 rows=998000 loops=1)
         ->  Subquery Scan on "*SELECT* 1"  (cost=0.00..61.00 rows=2550 width=8) (actual time=0.033..0.310 rows=1000 loops=1)
               ->  Seq Scan on p  (cost=0.00..35.50 rows=2550 width=4) (actual time=0.029..0.169 rows=1000 loops=1)
         ->  Subquery Scan on "*SELECT* 2"  (cost=0.00..213566.74 rows=40000 width=8) (actual time=423.472..589.881 rows=997000 loops=1)
               ->  Subquery Scan on q  (cost=0.00..213166.74 rows=40000 width=4) (actual time=423.472..549.365 rows=997000 loops=1)
                     ->  HashSetOp Except  (cost=0.00..212766.74 rows=40000 width=12) (actual time=423.471..507.091 rows=997000 loops=1)
                           ->  Append  (cost=0.00..180110.16 rows=6531315 width=12) (actual time=0.029..177.786 rows=997000 loops=1)
                                 ->  Subquery Scan on "*SELECT* 1_1"  (cost=0.00..146383.62 rows=6502500 width=12) (actual time=0.029..132.551 rows=997000 loops=1)
                                       ->  Nested Loop  (cost=0.00..81358.62 rows=6502500 width=8) (actual time=0.029..81.181 rows=997000 loops=1)
                                             ->  Seq Scan on s  (cost=0.00..35.50 rows=2550 width=4) (actual time=0.016..0.173 rows=997 loops=1)
                                             ->  Materialize  (cost=0.00..48.25 rows=2550 width=4) (actual time=0.000..0.027 rows=1000 loops=997)
                                                   ->  Seq Scan on p p_1  (cost=0.00..35.50 rows=2550 width=4) (actual time=0.006..0.096 rows=1000 loops=1)
                                 ->  Subquery Scan on "*SELECT* 2_1"  (cost=338.29..1069.96 rows=28815 width=12) (actual time=0.194..0.194 rows=0 loops=1)
                                       ->  Merge Join  (cost=338.29..781.81 rows=28815 width=8) (actual time=0.193..0.194 rows=0 loops=1)
                                             Merge Cond: (r.a = p_2.a)
                                             ->  Sort  (cost=158.51..164.16 rows=2260 width=8) (actual time=0.055..0.058 rows=100 loops=1)
                                                   Sort Key: r.a
                                                   Sort Method: quicksort  Memory: 29kB
                                                   ->  Seq Scan on r  (cost=0.00..32.60 rows=2260 width=8) (actual time=0.016..0.020 rows=100 loops=1)
                                             ->  Sort  (cost=179.78..186.16 rows=2550 width=4) (actual time=0.082..0.102 rows=995 loops=1)
                                                   Sort Key: p_2.a
                                                   Sort Method: quicksort  Memory: 71kB
                                                   ->  Seq Scan on p p_2  (cost=0.00..35.50 rows=2550 width=4) (actual time=0.004..0.035 rows=1000 loops=1)
 Planning Time: 0.548 ms
 Execution Time: 676.537 ms
(26 rows)
```

Results:

P/S size	R Size	Exec time Q7	Exec time Q8
10^3	10^3	77.736 ms	566.446 ms
10^2	10^4	323.528 ms	38.818 ms
10^3	10^2	11.023 ms	676.537 ms

A7 – Aniket Kale

Observation:
For equal sizes, Q7 was faster than Q8. For larger size of R, Q8 was much faster. In contrast, Q7 was much faster when the sizes were inversed. The differences in execution times in the 2 varying size example is drastic.

Question 5:

In the RA versions of the queries i.e. Q8, Q6, HashSetOp was prominently used. For the non-RA versions of the queries i.e. Q7, Q5 the queries used Nested Loop/Hash Anti join. However, as we can see the placement of P/S deeper in the query affects the performance hugely dependent on the size of the relations itself. As seen in Q7 and Q8 comparison, as R was deeper in the query, when it had a smaller size, the non-RA version performed better and vice versa.