☕ DATA 514 Section 6 Worksheet

Name: _____

Relational Algebra

Question 1 (4 points)

Fill in what the RA operators do.

σ =

π =

⋈ =

γ =

δ =

Question 2 (4 points)

Make this SQL query into RA (remember FJWGHOS)

```
SELECT R.b, T.c, max(T.a) AS T_max
  FROM Table_R AS R, Table_T AS T
 WHERE R.b = T.b
 GROUP BY R.b, T.c
HAVING max(T.a) > 99;
```

Question 3 (4 points)

Convert the following SQL queries into logical RA plans, given the following schemas:

Actor(aid, fname, lname, age)

ActsIn(aid, mid)

Movie(mid, name, budget, gross)

```
SELECT A.fname, A.lname, A.age
    FROM Actor AS A
    WHERE A.fname = ‘Patrick’
         AND A.lname = ‘Stewart’;
```

Question 4 (4 points)

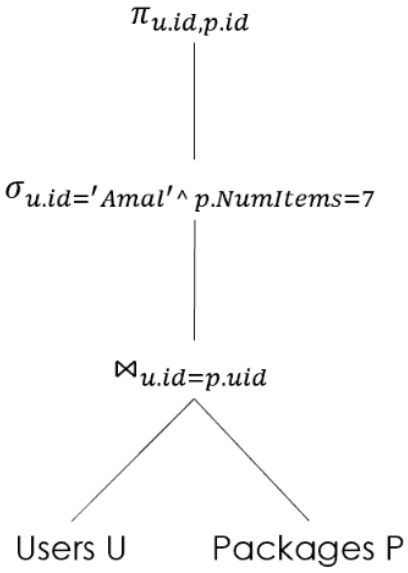
Convert the following SQL queries into logical RA plans, given the following schemas:
Actor(aid, fname, lname, age)
ActsIn(aid, mid)
Movie(mid, name, budget, gross)

```
SELECT M.name, COUNT(*) AS cnt
    FROM Actor AS A, ActsIn AS AI, Movie AS M
    WHERE A.aid = AI.aid AND M.mid = AI.mid
        AND A.age < 30
    GROUP BY M.mid, M.name
    HAVING COUNT(*) > 1;
```

Cardinality Estimation

Question 5 (4 points)

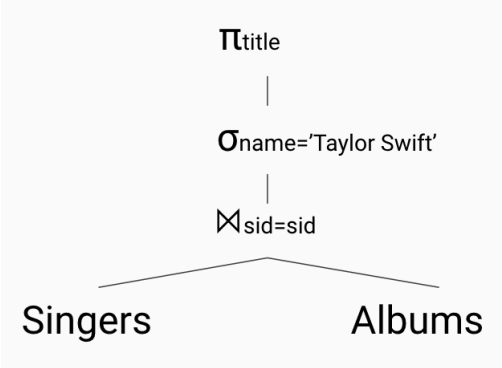
Consider the fact that Amazon has shipped several billion packages over the course of its >20y history and that it may surpass 10B packages by 2030. Assume that it tracks its packages and users using the following schema:
Packages(PackageID, UserID, DestAddress, NumItems)
Users(UserID, CreditCardNumber, Languages)
Now, consider the following RA tree:



You may notice how, although the PACKAGES table is very very large (10B!!), an individual user may have a very small number of rows. Generate a logically-equivalent tree which, ideally, takes advantage of this fact.

Question 6 (4 points)

Consider that many singers produce many albums over time. Assume the following schema:
Singers(sid, name, age, home_country) Albums(sid, title) Now, consider the following RA tree:



Consider the following statistics: - Singers Statistics: - T(Singers) = 5 - V(Singers, name) = 5 - Albums Statistics: - T(Albums) = 100
Rearrange the RA tree for a more advantageous cardinality estimate.
