



DATA 514 Section 6 Worksheet

Name: _____



Lecture Takeaways 6

1. Decompose into BCNF

2. Draw an RA tree for this query

```
SELECT MAX(UserID)
FROM Payroll
UNION
SELECT MAX(UserID)
FROM Regist
```

3. Draw an RA tree for this query

```
SELECT P.UserID,  
        COUNT(*)  
  FROM Payroll AS P,  
        Regist AS R  
 WHERE P.UserId = R.UserID  
GROUP BY P.UserID
```

4. An inner join is:

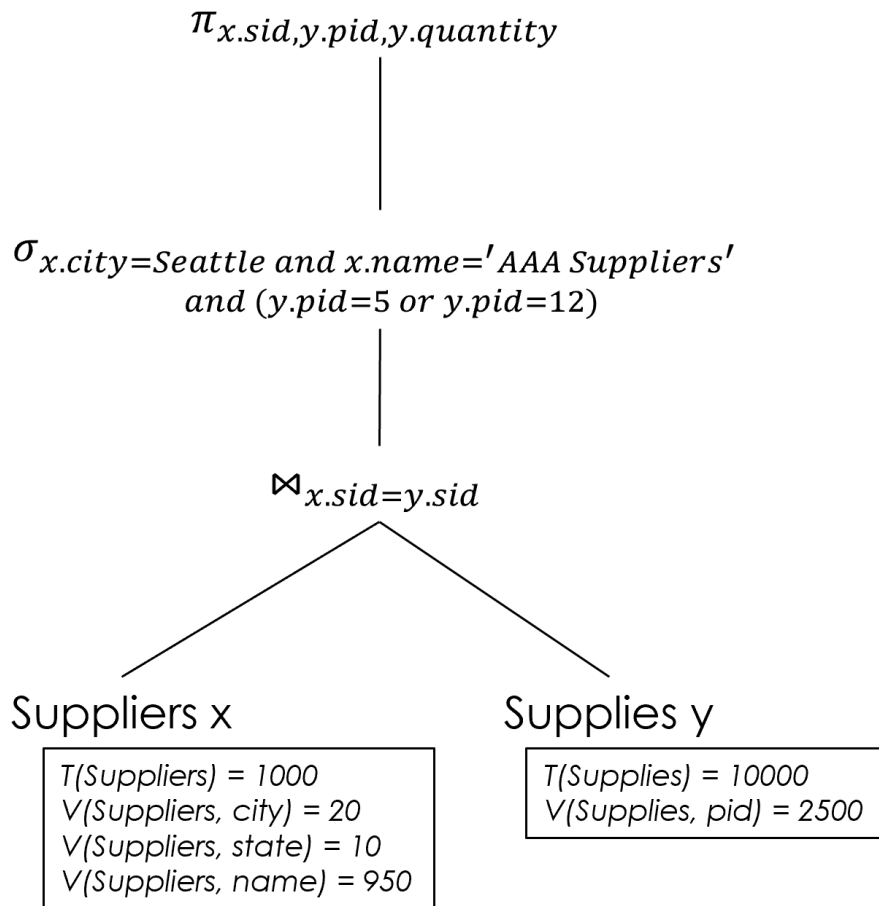
- Commutative only
- Associative only
- Both Commutative and Associative
- Neither Commutative nor Associative

5. Find a semantically-equivalent tree

6. Describe a uniformly-distributed dataset, or describe how one of our datasets does **NOT** have a uniform distribution

7. What is a general-purpose expression for the cardinality of a join between two tables on a primary key?

8. What is the estimated cardinality of this query?



Section Practice Problems

Relational Algebra

Question 1 (4 points)

Fill in what the RA operators do.

$\sigma =$

$\pi =$

$\bowtie =$

$\gamma =$

$\delta =$

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;
```

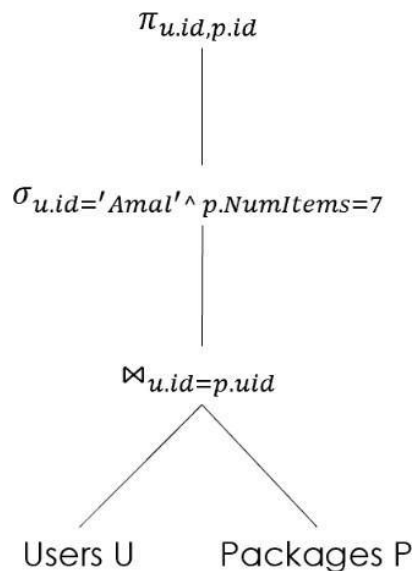
Cardinality Estimation

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)`

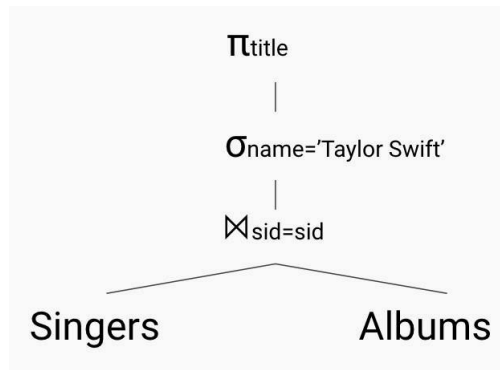
Now, consider the following RA tree:



You may notice how, although the PACKAGES table is very very large (10B!!), an individual user may have a very small number of rows. Use the space above to 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: - $T(\text{Singers}) = 5$ - $V(\text{Singers}, \text{name}) = 5$ - Albums Statistics: - $T(\text{Albums}) = 100$

Rearrange the RA tree for a more advantageous cardinality estimate.
