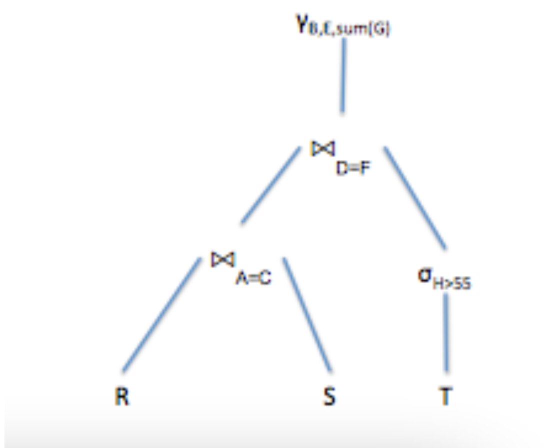
Relational Algebra, Relational Calculus, Datalog

Past Midterm problems

1) Schema: R(A, B), S(C, D, E), T(F, G); Spring, 2011

Write Relational Algebra Plan for the SQL query below. Your answer should be a tree representing the relational algebra plan.

```
SELECT R.B, S.E, sum(T.G)
FROM R, S, T
WHERE R.A = S.C and S.D = T.F and T.H > 55
GROUP BY R.B, S.E
```

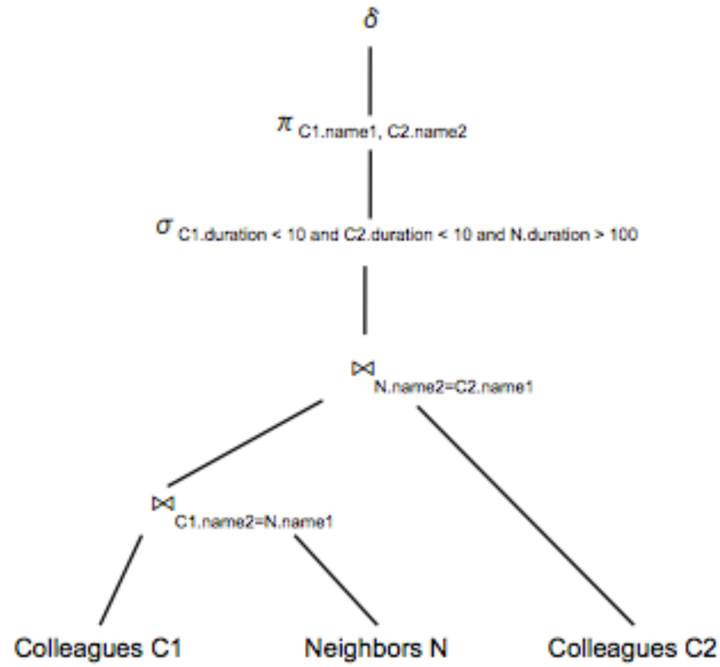


2) Consider the following database schema (Fall 2011)

Neighbors(name1,name2,duration), Colleagues(name1,name2,duration)

(a) Write a Relational Algebra Plan for the SQL query below. Your answer can be in the form of an expression or a tree, whichever you prefer:

```
SELECT DISTINCT C1.name1, C2.name2
FROM Colleagues C1, Neighbors N, Colleagues C2
WHERE C1.name2 = N.name1
AND N.name2 = C2.name1
AND C1.duration < 10
AND C2.duration < 10
AND N.duration > 100
```



(b) Write a Datalog query that returns all neighbors who do not have any colleagues in common.

Solution: NonAnswers(n1, n2):- Neighbors(n1, n2, -), Colleagues(n1, c, -), Colleagues(n2, c, -)
A(n1, n2):- Neighbors(n1, n2, -), NOT NonAnswer(n1,n2)

(c) Indicate if the following relational calculus queries are correct or not (true or false). Note: This is not meant to be a tricky question. Errors, if any, should be reasonably obvious. You do NOT need to correct wrong queries:

- $A(x) = \exists y \exists z\ Neighbor(x, y, -)\ \wedge Colleagues(y, z, -)$ **TRUE**
- $A(x) = Neighbor(x, -, -)\ \wedge\ (\forall y\ Neighbors(x, y, -)\ \wedge Colleagues(x, y, -))$ **FALSE**
- $A(x) = Neighbors(x, -, -) \wedge (\forall y\ Neighbors(x, y, -) \rightarrow \exists z Colleagues(y, z, -))$ **TRUE**

Extra Stuffs!

Indices

For each statement below, indicate whether it is true of false.

i. An index may help a select-from-where SQL query run faster, or may not affect its running time, but it can never make a query run slower. **TRUE**

ii. ii. An index may help an update (insert, delete, or update) SQL query run faster, or may not affect its running time, but it can never make a query run slower. **FALSE**

iii. Consider a selection operation $\sigma_{price>90 \wedge price<100}(Product)$ Using an unclustered index on price will make the query at least as fast as scanning the entire table Product. **FALSE**

iv. Consider a selection operation $\sigma_{price>90 \wedge price<100}(Product)$. Using a clustered index on price will make the query at least as fast as scanning the entire table Product. **TRUE**

v. A large table Product(pid, name, price) is queried intensively and is never updated. Then we should create three clustered indexes, on Product(pid), Product(name), and Product(price). **FALSE**

Relational Calculus Domain Independence

Consider the relations R(x, y) and S(x), show the answer of each relational calculus query.

1) R = {(10, 10), (10, 20), (20, 10), (30, 20)}, S = {10, 20}

2) R = {(10, 10), (10, 20), (20, 10), (30, 20)}, S = {10}

3) R = {(10, 10), (10, 20), (20, 10), (30, 20)}, S = {}

For each of them, find $Q(x){:}-\ \forall y(S(y) \rightarrow R(x, y))$, is this relational calculus domain independent?

1) Ans = { 10 }
2) Ans = { 10, 20 }
3) Ans = { 10, 20, 30 }

No, the relational calculus is domain dependent.

Cheat sheet for relational algebra

Name	Symbol
Selection	σ
Projection	π
Join	$\bowtie$
Group By	γ
Set Difference	$-$
Duplicate Elimination	δ