CS 4750 – Assignment 4

Geoff Hicks (gdh8jh), Sammy Lahrime (sl8rn), Sam Book (slb8pm)

Part 1: Relational Algebra

```
advisor (s_ID, i_ID)      -- s_ID references student(ID), i_ID references instructor(ID)
classroom (building, room_number, capacity)
course (course_id, title, dept_name, credits)
department (dept_name, building, budget)
instructor (ID, name, dept_name, salary)
prereq (course_id, prereq_id)      -- prereq_id references prereq(course_id)
section (course_id, sec_id, semester, year, building, room_number, time_slot_id)
student (ID, name, dept_name, tot_cred)
takes (ID, course_id, sec_id, semester, year, grade)      -- ID references student(ID)
teaches (ID, course_id, sec_id, semester, year)       -- ID references instructor(ID)
time_slot (time_slot_id, day, start_time, end_time)
```

1.1 Find the IDs and names of all students who have not taken any course offering before Spring 2009.

$$(\pi_{name,\ ID}(student)) \ - (\pi_{name,\ ID}(\sigma_{year < 2009}(student \bowtie takes)))$$

1.2 Find the course sections taught by more than one instructor without using any aggregate functions

$$\rho_{T1(ID,course\_id,\text{sec }\_id,semester,year)}(teaches)$$

$$\rho_{T2(ID,course\_id,\text{sec }\_id,semester,year)}(teaches)$$

$$\pi_{sec\_id}(\sigma_{T1.id<>T2.id}(T1 \bowtie T2))$$

```
Product (maker, model, type)     -- maker is foreign key to Manufacturer(maker)
                                 -- type is stored as pc, laptop, printer
PC (model, speed, ram, hd, price)
Laptop (model, speed, ram, hd, screen, price)
Printer (model, color, type, price)  -- color: true for color printer, false for black/white printer
```

2.1 Find the model number and price of all products (of any type) made by manufacturer 'BBB'

$$\pi_{model,\ price}(\sigma_{maker="BBB"}(PC \bowtie Product))$$

$$\cup$$

$$\pi_{model,\ price}(\sigma_{maker="BBB"}(Laptop \bowtie Product))$$

$$\cup$$

$$\pi_{model,\ price}(\sigma_{maker="BBB"}(Printer \bowtie Product))$$

2.2 Find those hard-disk (hd) sizes that occur in two or more PC's.

$$\rho_{PC1(model,speed,ram,hd,price)}(PC)$$

$$\rho_{PC2(model,speed,ram,hd,price)}(PC)$$

$$\pi_{hd}(\sigma_{PC1.hd=PC2.hd\ AND\ PC1.model<>PC2.model}(PC1 \bowtie PC2))$$

```
Supplier (sid, sname, address)
Part (pid, pname, color)
Catalog (sid, pid, cost)    -- which supplier suppies which part and at which price (cost)
```

3.1 Find the names of suppliers who supply some red part.

$$\pi_{sname}(\sigma_{color="red"}(Part \bowtie Catalog \bowtie Supplier)$$

3.2 Find the IDs of suppliers who supply only red parts

$$\pi_{sid}(Supplier) - \pi_{sid}(Catalog \bowtie (\sigma_{color <>"red"}(Part)))$$

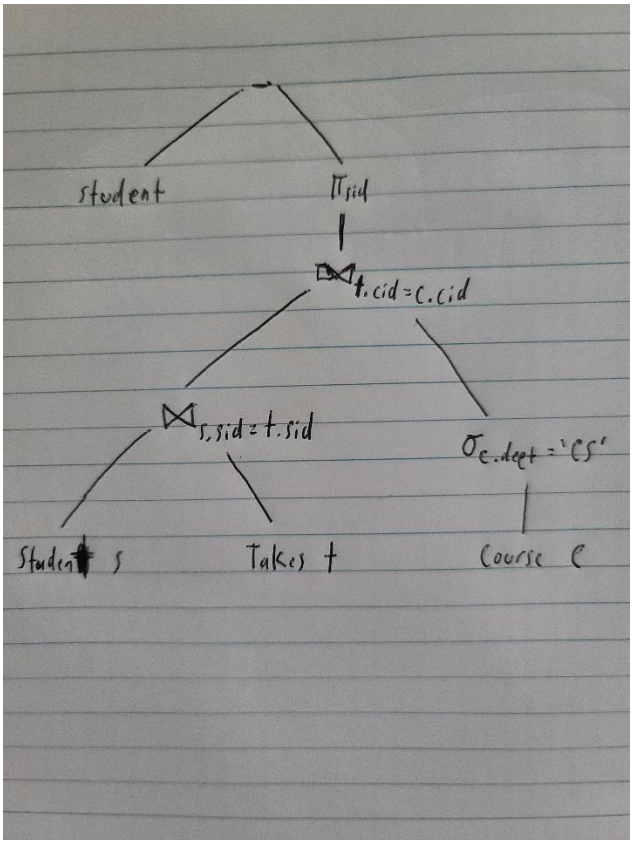
3.3 Find the IDs of suppliers who supply all parts with cost < 20.00

$$\pi_{sid,pid}($$
$$(Part \bowtie Supplier)$$
$$\bowtie$$
$$((\pi_{sid}(Part \bowtie Supplier))$$
$$-$$
$$(\pi_{sid}(\sigma_{cost > 20}(Part \bowtie Supplier))))$$
$$)$$

Part 2: Cost Estimation

1.1 Write a logical plan (i.e., an RA tree) for the following query.

```
SELECT s.sid
FROM   Student s
WHERE  NOT EXISTS
       (SELECT t.sid
        FROM   Takes t, Course c
        WHERE   s.sid = t.sid
          AND   t.cid = c.cid
          AND   c.dept = 'CS');
```



```
T(R) = 10^5              // total number of rows in relation R
T(S) = 6 * 10^6          // total number of rows in relation S
T(T) = 5 * 10^4          // total number of rows in relation T
V(R,A) = 5 * 10^4        // number of distinct values of attribute A in relation R
V(R,B) = 3 * 10^3        // number of distinct values of attribute B in relation R
V(S,B) = 3 * 10^3        // number of distinct values of attribute B in relation S
V(S,C) = 2 * 10^4        // number of distinct values of attribute C in relation S
V(T,C) = 2 * 10^4        // number of distinct values of attribute C in relation T
V(T,D) = 10^4            // number of distinct values of attribute D in relation T
```

2.1 Estimate the number of tuples returned by σA=4750(R). Show your steps.

1. We know that in regardes to selectivity ($\sigma_{A=c}$(R)), we have the form $\frac{1}{V(R,A)}$

2. Thus, we will get $10^5 * \frac{1}{5*10^4}$ which will equal 2

2.2 Estimate the number of tuples returned by the following query. Show your steps.

```
SELECT R.A, S.B, S.C
FROM   R NATURAL JOIN S
```

$\frac{T(R)*T(S)}{\max\{V(R,B),V(S,B)\}}$ -> $\frac{(10^5)*(6*10^6)}{\max\{3*10^3\ ,\ 3*10^3\}}$ -> $\frac{(10^5)*(6*10^6)}{3*10^3}$ -> 200000000

2.3 Estimate the number of tuples returned by the following query. Show your steps.

```
SELECT *
FROM   R, S, T
WHERE  R.A = 4750 and R.B = S.B and S.C = T.C and T.D = 1102
```

1. $T(R) * T(S) * T(T)$ -> $10^5 * (6 * 10^6) * (5 * 10^4)$ -> $3 * 10^{16}$

2. $(3 * 10^{16}) * \left(\frac{1}{V(R,A)}\right) * \left(\frac{1}{\max\{V(R,B),V(S,B)\}}\right) * \left(\frac{1}{\max\{V(S,C),V(T,C)\}}\right) * \left(\frac{1}{V(T,D)}\right)$

3. $(3 * 10^{16}) * \left(\frac{1}{5*10^4}\right) * \left(\frac{1}{3*10^3}\right) * \left(\frac{1}{2*10^4}\right) * \left(\frac{1}{10^4}\right)$ -> 1