CS34800 Fall 2016 Assignment 2

Due 11:59pm 23 September, 2016

Given the following tables (**Events** and **Rooms**), answer questions A and B.

Events

Name	*Date*	*Location*
NextGear Capital	Sep 22	LWSN 1142
Film: In His Own Home	Sep 23	WTHR 172
OUTfest 2016	Sep 17	Downtown Lafayette
TechPoint	Sep 20	LWSN 1142
CS Outstanding Alumni	Sep 23	LWSN 3102
Ken Burns The American Experience	Sep 22	Elliott Hall
Big Mama Speaks	Sep 22	Black Cultural Center

Rooms

Location	*Capacity*	*Address*
LWSN 1142	104	305 N University St
WTHR 172	155	560 Oval drive
LWSN 3102	44	305 N University St
Downtown Lafayette	10000	400 Main Street
Elliot Hall	6005	712 3rd St
Black Cultural Center	36	1100 3rd St

A. Give the results of the following relational algebra expressions:

(a) $\Pi_{Name,Location}(\sigma_{Date=Sep22}(Events))$

(b) $\Pi_{Address,Date}(Events \bowtie_{E.Location=R.Location} Rooms)$

(c) $\gamma_{Date,count(*)}(Events)$

(d) $\gamma_{Address,count(*)}(Events \bowtie_{E.Location=R.Location} Rooms)$

(For those who find the typography hard to follow, $\bowtie$ is join, and γ is the group by operator.)

B. Eight queries are given below. However, there are really four pairs of queries - for each query, there is another that gives exactly the same result. For example, queries (a) and (b) do the same thing. Please identify the other three pairs.

(Note that this could be a Relational Algebra expression matching another Relational Algebra expression, a SQL query matching another SQL query, or a Relational Algebra expression matching an SQL query.)

(a) $\sigma_{Capacity=10000}(Rooms)$

(b) $\sigma_{Location='Downtown\ Lafayette'}(Rooms)$

(c)
```
SELECT DISTINCT Address
FROM Rooms
WHERE Capacity < 1000
```

(d) $\sigma_{Events.Location=Rooms.Location}(Events \times (\sigma_{Capacity<1000}(Rooms)))$

(e)
```
SELECT Address
FROM (SELECT MAX(Capacity), Address
FROM Rooms GROUP BY Address);
```

(f) $\sigma_{Capacity<1000}(Events \bowtie_{E.Location=R.location} Rooms)$

(g)
```
SELECT DISTINCT Address
FROM Rooms
WHERE Capacity != 44 or Capacity > 30;
```

(h) $\Pi_{Address}(\sigma_{Capacity<1000}(Rooms))$

C. Given a table **T** with attributes a and b (where both a and b are integers), explain what the following relational algebra expression does, and also show the result:

$$\sigma_{Tleft.a>T.max(a)\wedge Tleft.b=T.b}(\rho_{Tleft}(T) \times (\gamma_{b,max(a)}(T)))$$

D. Given the following tables (with which you should be quite familiar):

```
MUSICIAN (MNO, NAME, ADDRESS, PHONE)
ALBUM (ANO, ATITLE, COPYRIGHT_DATE, MNO, COPY_SOLD)
SONG (SNO, STITLE, ANO, LYRICIST)
PLAY (MNO, INSTRUMENT)
PERFORM (MNO, SNO, INSTRUMENT)
```

- The MUSICIAN relation contains information about the musicians in the system. The MNO attribute uniquely identifies each musician.
- The ALBUM relation contains information about the albums. The ANO attribute acts as the unique identifier of an album; MNO refers to a musician already existing in the MUSICIAN relation and acts as the producer of the album. Other information is: title, copyright date, and the number of copies that have been sold. An album contains many songs. An album may have many songs performed by various musicians, but it has only one musician as its producer.
- The SONG relation has SNO as the unique identifier. ANO refers to the album in which the song appears. LYRICIST is the name of the lyricist who wrote the lyrics of the song.

- The PLAY relation contains information about the instruments that a musician can play.
- The PERFORM relation contains information about the instrument that a musician plays in each song. Note that a musician can play many instruments, but in a particular song, he/she plays only one instrument.

Write relational algebra expressions for the following questions (many of which should also be familiar):

(a) Find the names of musicians who cannot play 'Piano'.

(b) Find the names of musicians who only use 'Piano' or 'Guitar' to perform.

(c) Find the phone numbers of all musicians who had used 'Guitar' to perform.

(d) Find the ANOs of the albums which have a song performed by 'Guitar'.

(e) Find the ANOs of the albums which only have songs performed by 'Guitar'.

(f) Find the MNO of the musician who produces the largest number of albums.

(g) Find the names of the musicians who can play all instruments.

(h) Find the names of the musicians who can play only one instrument.

(i) Find the names of the musicians who have never performed a song produced by other musicians.

(j) Find the ANOs of the albums that have a song performed by other musicians (not the producer).

(k) Find the names of the musicians who have performed for albums that have a copyright before '03-Sep-1999'.

(l) Find the name of the musician who has the album with the maximum number of sold copies.

(m) Find the names of (distinct) musicians who can play 2 or more instruments.

(n) In Western classical music, there is a lot of chamber music written for *string quartets* (e.g., the Verona String Quartet (`https://www.purdue.edu/convocations/event/verona-string-quartet-daniel-hsu-piano/`), ETHEL (`http://ethelcentral.org/`).) A string quartet consists of exactly four instruments: two 'violin's, one 'viola', and one 'cello'. Give relational algebra to find all of the songs by string quartets.

Hint: You need songs that contain the exact number of each of the four instruments, and don't contain any other instruments. How can you find *don't contain any other instruments* without knowing in advance what other instruments there might be? There are multiple ways to do this (some easier than others), but you'll need to think about it.