Indian Statistical Institute

PGDBA, First Year, Mid-Sem of First Semester Examination, 2020-21

Fundamentals of Database Systems

Full Marks: 30 Date: 30-12-2020 Time: 2 Hours

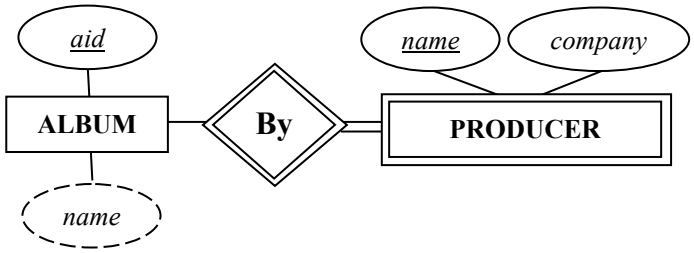
Answer any *three* of the following questions

$3 \times 10 = 30$

1. (a) Prove or disprove the following statement: The relational algebra operation $R_1 - R_2$ return no less than the number of tuples returned by $R_1 \div \pi_X(R_2)$, for any arbitrary pair of relations R_1 and R_2. Note that $A(R_1) = A(R_2)$ and $X \subset A(R_2)$, where $A(R)$ represents the set of attributes in a relation R.
(b) Consider the two relations $R_1(V, W)$ and $R_2(X, Y, Z)$ having the number of tuples $t1 > 0$ and $t2 > 0$ tuples, respectively. Without making any assumptions on the keys, find out the minimum and maximum possible number of tuples that may appear in the relation resulting from the expression $R_1 \ltimes R_2$. Note that the notion $\ltimes$ denotes a semijoin, which is operationally close to natural join with the only exception that attributes in the first relation are only returned in the result.

6+4

2. (a) Note that complex constraints in E-R data model are labeled along the edges between an entity set and a binary relationship set. Comment about the validity of the following complex constraints with appropriate justifications.
(i) 1 . . *
(ii) * . . *
(iii) 0 . . 0
(b) Identify errors in the following E-R diagram representing details about music albums and their producers. The *aid* and *name* in the relation ALBUM denote the album ID and name of the album, respectively. The *name* and *company* in the relation PRODUCER denote the name of the producer and the respective company sponsoring the album, respectively.



(2+2+2)+4

3. (a) Let there be a relation $R(F, U, N)$, in which the attribute F is numeric. Write SQL queries that can find out the average value of the attribute F without using the avg() function.
(b) Suppose there is a single table that can keep the details of winners in Grand Slam women's singles champions in tennis. The schema of the table is as follows.

GRAND = <*year* : integer, *event* : string, *name* : string, *nationality* : string>

The primary keys are underlined in the above table. Note that the *event* represents the four events that are the part of Grand Slam championship, namely Wimbledon, Australian Open, French Open, and US Open. Create a trigger that can distribute the tuples being entered into this table (retaining the original tuples) into 4 different tables (already existing) keeping the details about different events. The schemas of these tables are as follows.

WIMBLEDON = <*year* : integer, *name* : string, *nationality* : string>
AUSTRALIAN = <*year* : integer, *name* : string, *nationality* : string>
FRENCH = <*year* : integer, *name* : string, *nationality* : string>
US = <*year* : integer, *name* : string, *nationality* : string>

3+7

4. (a) Let there be a pair of relations $R1$ and $R2$, with uniformly distributed values over the attributes, both having a combination of two attributes serving as their primary keys. If $R1$ and $R2$ share only a single common attribute, say A^*, estimate the sizes of the following queries. Assume that the number of tuples in $R1$ and $R2$ are $t1$ and $t2$, respectively. Further consider that there are $d1$ and $d2$ distinct entries of the attribute A^* in $R1$ and $R2$, respectively.

(i) $R_1 \bowtie R_2 \cap R_1 \bowtie R_2$.

(ii) $R_1 \bowtie R_2$.

(iii) $R_1 - (R_1 - R_2)$.

(b) Give a counterexample to disprove the equivalence $\sigma_\theta (R_1 \cup R_2) \equiv \sigma_\theta (R_1) \cup R_2$, where θ is a predicate.

(3+3+3)+1