CO526 Databases Course Work 1: SQL

Due in 12noon 15th February 2019

Background Material

The **family_history** database (available on both SQLServer and Postgres in DoC) contains a table person, the schema and some data from which are listed below. Note that dob is the date of birth, and dod is the date of death. Any optional (nullable) columns are shown with a question mark against the column name.

person						
<u>name</u>	gender	dob	dod?	father?	mother?	born_in
Alice	F	1885-02-25	1969-12-05	null	null	Windsor
Andrew	M	1960-02-19	null	Philip	Elizabeth II	London
Andrew of Greece	M	1882-02-02	1944-12-03	George I of Greece	null	Athens
Anne (Princess)	F	1950-08-15	null	Philip	Elizabeth II	London
Charles	M	1948-11-14	null	Philip	Elizabeth II	London
			⋮			

$\mathsf{person(father)} \overset{\mathsf{fk}}{\Rightarrow} \mathsf{person(name)}$
$\mathsf{person(mother)} \overset{\mathsf{fk}}{\Rightarrow} \mathsf{person(name)}$

Submission

To gain full marks, answers to the following questions should make full use of ANSI SQL commands to write compact and efficient queries, and be laid out such that structure of the query is clear. The queries must also run correctly on the Postgres version of the database, and be submitted electronically to CATE as single batch file db_2019_cw1.sql by the coursework deadline. A template version of the file is available on CATE for download. The queries in the file must be given in the order of the questions below, and be separated by semi-colons.

To test your answer against the Postgres version of the database, you should run the command:

```
psql -h db.doc.ic.ac.uk -d family_history -U lab -W -f db_2019_cw1.sql
```

Note that 60% of the marks will be awarded for correctness, and 40% of the marks for style, including how consise the queries are, appropriate use of indentation, use of Capital letters for keywords, and exprsssing join conditions by use of JOIN statements in the FROM clause as opposed to using equals in the WHERE clause.

Questions

1. Translate the following RA query (where $\mathsf{person_a}$ and $\mathsf{person_b}$ are aliases for person) into an equivalent query expressed in SQL:

$$\pi_{\mathsf{person_b.name,person_a.dod}}\, \sigma_{\mathsf{person_a.name=person_b.mother} \land \mathsf{isNotNull(person_a.dod)}}(\mathsf{person_a} \times \mathsf{person_b})$$

2. Write an SQL query returning the scheme (name) ordered by name that lists men not known to be fathers.
3. Write an SQL query that returns the scheme (name) ordered by name containing the name of mothers who have had every gender of baby that appears in the database. You must not write the query to assume that gender is limited to 'M' and 'F'.
4. Write a query that returns the scheme (name,father,mother) ordered by name that lists the name of people who are the first born of their known full siblings. A full sibling is defined as a person sharing the same father and mother. Maximum marks will be given only to answers that use ALL or SOME to answer the question.
5. Write an SQL query that returns the scheme (name,popularity) ordered by popularity,name listing first names and the number of occurances of first names. A first name is taken to mean the first word appearing the name column. The most popular first name must be listed first, and the list must exclude any first name that appears only once.
6. Write an SQL query that returns the scheme (name,forties,fifties,sixties) ordered by name listing one row for each person in the database whom has had at least two children, and for each such person, gives three columns forties, fifties and sixties containing the number of that person's children born in those 20th century decades.
7. Write an SQL query returning the scheme (father,mother,child,born) that lists known fathers and mothers of children, with born being the number of the child of the parents (*i.e.* returning 1 for the first born, 2 for the second born, *etc*). The result should be ordered by father,mother,born
8. Write an SQL query that returns the scheme (father,mother,male) ordered by father,mother that lists all pairs of known parents with the percentage (as a whole number) of their children that are male.