CO526 Databases Course Work 1: SQL

Due in noon Friday 16th February 2018

The tables below gives details of a **family_history** database. In the person table, people are identified by their name, and always have their gender, date of birth (dob) and place of birth (born_in) recorded. In addition, each person may optionally have recorded the name of their father, and the name of their mother. If the person has died, then the date of death dod must be present. Note that only a fragment of the data held in the database is listed below.

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
			⋮			

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

In addition, there is a table monarch which contains the English head of state — normally a monarch (*i.e.* a King or Queen) — where the house of each monarch indicates which royal house the monarch belongs to, accession indicates the date the person came to the throne, and coronation the date any coronation of the monarch. If null appears in coronation then the person had no coronation. Each monarch remains head of state until the succession of the next monarch. Note that the value of null appearing in house indicates *not* a King or Queen, but a head of state (for example Oliver Cromwell) who filled the role of Protector during the Commonwealth period in the 17th Centuary.

monarch			
<u>name</u>	house?	accession	coronation?
James I	Stuart	1603-03-24	1603-07-25
Charles I	Stuart	1625-03-27	1626-02-02
Oliver Cromwell	null	1649-01-30	null
Richard Cromwell	null	1658-09-03	null
Charles II	Stuart	1659-05-25	1626-02-02
James II	Stuart	1685-02-06	1685-04-23
		⋮	

monarch(name) $\overset{fk}{\Rightarrow}$ person(name)

Finally, there is a table prime_minister, recording the party the person led whilst Prime Minister, and the date of entry into office. A person remains Prime Minister until the date of entry to office of the next Prime Minister. Note a person may have more than one period in office.

prime_minister		
<u>name</u>	party	<u>entry</u>
David Cameron	Conservative	2010-05-11
Gordon Brown	Labour	2007-06-27
Tony Blair	Labour	1997-05-02
John Major	Conservative	1990-11-28
Margaret Thatcher	Conservative	1979-05-04
James Callaghan	Labour	1976-04-05
Harold Wilson	Labour	1974-03-04
Edward Heath	Conservative	1970-06-19
	⋮	

prime_minister(name) $\overset{fk}{\Rightarrow}$ 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_2018_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, with the comment present in the template file left unchanged, and the query terminated by a single semi-colon.

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_2018_cw1.sql
```

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

Questions

The first four questions test knowledge of SQL as an implementation of the Relational Algebra, and the last four questions test knowledge of SQL as a Programming Language.

Style marks to apply once over the whole exercise
ⒶLoose 4 marks if more than one instance of using commas to sepate tables in FROM clause, rather than using JOIN. If no use or only one use is made of JOIN, then loose 8 marks.
ⒷLoose 8 marks if all keywords in lower case.
ⒸLoose 8 marks if no use of indentation is made, loose 4 marks if inconsistent use of indentation is made
ⒹLoose 4 marks for any solution that uses temporary tables when the temporary table is used only once. Using temporary tables prevents query optimisation between the temporary table and the main query. CTE (WITH) statements can be used.
Style marks to apply once per query
Ⓔ Loose 3 marks each time an unnessary subquery is used (ie where instead the same result could be achieved by joining inside the same SELECT, or where the result of a subquery does not require processing by the outer query).
Correctness marks to apply once per query
ⒻLoose 5 marks for each query that does not run due to syntax errors
ⒼLoose 2 marks for each question that omits an ORDER BY
ⒽLoose 3 marks for each missing WHERE or HAVING condition
ⒾLoose 3 marks for each question failing to return columns in order requested by question, or failing to return the requested columns.

1. Write an SQL query that returns the scheme (name,father,mother) ordered by name containing the name of all people known to have died before both their father and mother, together with the name of the mother and the name of the father.

```
SELECT  person.name,
        father.name AS father,
        mother.name AS mother
FROM    person
        JOIN person AS father
        ON   person.father=father.name
        JOIN person AS mother
        ON   person.mother=mother.name
WHERE   person.dod<father.dod
AND     person.dod<mother.dod
ORDER BY person.name
```

10

2. Write an SQL query returning the scheme (name) ordered by name that lists all people that have either been a King, Queen or Prime Minister.

```
SELECT name
FROM   monarch
WHERE  house IS NOT NULL
UNION
SELECT name
FROM   prime_minister
ORDER BY name
```

ⒿLoose 4 marks if JOIN used instead of UNION
ⓀLoose 2 marks for using DISTINCT in the SELECT statements (since UNION will eliminate the duplicates).
Ⓔ Applies here if this UNION is put inside a SELECT to get rid of duplicates (unncessary since UNION does that by default) or to join/check membership of person (unncessary because of the foreign keys).

10

3. A King or Queen is said to abdicate if their reign ceases before their death. Write an an SQL query returning the scheme (name) ordered by name that lists the name of all Kings or Queens that have abdicated

```
SELECT   monarch.name
FROM     monarch
         JOIN person
         ON   monarch.name=person.name
WHERE    EXISTS (SELECT *
                 FROM   monarch AS next_monarch
                 WHERE  next_monarch.accession>monarch.accession
                 AND    person.dod>next_monarch.accession)
AND      monarch.house IS NOT NULL
ORDER BY monarch.name
```

12

ⓁLoose 3 marks for listing Mary II who reigned jointly with William III, and died before William III (and thus did not abdicate, despite the next monarch being after her death).

4. Write a query that returns the scheme (house,name,accession) ordered by accession that lists house and name of monarchs who were the first of a house to accede to the throne. Maximum marks will be given only to answers that use either the ALL or SOME operators.

```
SELECT  monarch.house,
        monarch.name,
        monarch.accession
FROM    monarch
WHERE   monarch.accession<=ALL (SELECT later_monarch.accession
                                FROM   monarch AS later_monarch
                                WHERE  monarch.house=later_monarch.house)
AND     monarch.house IS NOT NULL
ORDER BY accession
```

12

5. Write an SQL query that returns the scheme (first_name,popularity) ordered in descending order of popularity, and then alphabetical order of first_name. Your answer should also exclude first names that only occur once in the database. A first name is taken to mean the first word appearing the name column of person.

```
SELECT    first_name,
          COUNT(first_name) AS popularity
FROM      (SELECT CASE
                  WHEN POSITION(' ' IN name)=0 THEN name
                  ELSE SUBSTRING(name FROM 1 FOR POSITION(' ' IN name)-1)
                  END AS first_name
           FROM   person) AS person
GROUP BY  first_name
HAVING    COUNT(first_name)>1
ORDER BY  popularity DESC, first_name;
```

ⓂLoose 5 marks for not using standard SQL string functions but instead using non standard functions such as CHARINDEX.
ⓃLoose 2 marks for getting the string breaks wrong, so that letters are missing from names, or spaces added to the end of names.

12

6. Write an SQL query that returns the scheme (house,seventeenth,eighteenth,nineteenth,twentieth) ordered by house listing the number of monarchs of each royal house that acceded to the throne in the 17th, 18th, 19th and 20th centuries.

```
SELECT house,
       COUNT(CASE WHEN EXTRACT(YEAR FROM accession) BETWEEN 1600 AND 1699
                  THEN accession ELSE null END) AS seventeenth,
       COUNT(CASE WHEN EXTRACT(YEAR FROM accession) BETWEEN 1700 AND 1799
                  THEN accession ELSE null END) AS eighteenth,
       COUNT(CASE WHEN EXTRACT(YEAR FROM accession) BETWEEN 1800 AND 1899
                  THEN accession ELSE null END) AS nineteenth,
       COUNT(CASE WHEN EXTRACT(YEAR FROM accession) BETWEEN 1900 AND 1999
                  THEN accession ELSE null END) AS twentieth
FROM   monarch
WHERE  house IS NOT NULL
GROUP BY house
ORDER BY house
```

It is acceptable to use comparison operators between dates instead of the EXTRACT function to get the year.
ⓄLoose 2 marks for converting the date to a string and using string comparisons (or converting back to numeric format for comparison).
ⓅLoose 5 marks if query using UNION given. 12

7. Write an SQL query returning the scheme (father,child,born) ordered by father,born that lists as father the name of all men in the database, together with the name of each child, with born being the number of the child of the father (*i.e.* returning 1 for the first born, 2 for the second born, *etc*). For men with no children, the man should be listed with null for both child and born.

```
SELECT person.name AS father,
       child.name AS child,
       CASE
       WHEN child.name IS NOT NULL
       THEN RANK() OVER (PARTITION BY child.father ORDER by child.dob)
       ELSE null
       END AS born
FROM   person
       LEFT JOIN person AS child
       ON person.name=child.father
WHERE  person.gender='M'
ORDER BY person.name,
       born
```

Note that the question does not specify a sort order where the rank is tied, and hence George IV and William IV may appear in either order as children of George III.
ⓆLoose 4 marks for ranking joint birth dates lower (ie if joint seconds are listed as joint thirds).
ⓇLoose 3 marks for including women (!) as fathers. 16

8. Write an SQL query that returns the scheme (monarch,prime_minister), ordered by monarch and prime minister, that lists prime ministers that held office during the reign of the monarch.

```
SELECT DISTINCT monarch.name AS monarch,
        prime_minister.name AS prime_minister
FROM    monarch CROSS JOIN prime_minister
WHERE   -- No later prime minister coming before the monarch acceded
        NOT EXISTS (SELECT *
                    FROM    prime_minister AS later_prime_minister
                    WHERE   later_prime_minister.entry<=monarch.accession
                    AND     later_prime_minister.entry>prime_minister.entry)
        -- No monarch before this prime ministers entry
AND     NOT EXISTS (SELECT *
                    FROM    monarch AS later_monarch
                    WHERE   prime_minister.entry>=later_monarch.accession
                    AND     later_monarch.accession>monarch.accession)
AND     monarch.house IS NOT NULL
ORDER BY monarch,
        prime_minister
```

ⓈLoose 4 for exluding prime ministers that servered over the start and/or end of a reign, and hence for example missing Baldwin serving under Edward VIII.

ⓉLoose 3 for excluding the current monarch (and thus having no results for Elizabeth II) 16

Note that the NOT EXISTS may be replaced by similar expressions comparing the outer prime minister or monarch with the subquery using ALL.

TOTAL MARKS:100