PAPER C130

INTRODUCTION TO DATABASES

Monday 4 May 2020, 11:00
Duration: 80 minutes
Post-processing time: 30 minutes
Answer TWO questions

While this time-limited remote assessment has not been designed to be open book, in the present circumstances it is being run as an open-book examination. We have worked hard to create exams that assesses synthesis of knowledge rather than factual recall. Thus, access to the internet, notes or other sources of factual information in the time provided will not be helpful and may well limit your time to successfully synthesise the answers required.
Where individual questions rely more on factual recall and may therefore be less discriminatory in an open book context, we may compare the performance on these questions to similar style questions in previous years and we may scale or ignore the marks associated with such questions or parts of the questions. In all examinations we will analyse exam performance against previous performance and against data from previous years and use an evidence-based approach to maintain a fair and robust examination. As with all exams, the best strategy is to read the question carefully and answer as fully as possible, taking account of the time and number of marks available.

Paper contains 2 questions

Several parts of the following questions make use of the **mondial** relational database, a fragment of which is listed below. It contains information about countries, and the membership of countries in organisations.

For each organization, there is a record of the city and country in which that organisation is based, which if not present indicates the organisation has no official base.

The is_member table records the relationship of coutries to organisations, where countries may either be a full member or an observer to the organisation.

The borders table records which countries share a land border, and the length of that border. Note that each pair of neighbouring countries appears only once in borders.

organization

abbreviation	city?	country?	established?
AL	Cairo	ET	1945-03-22
C	London	GB	1931-12-31
CERN	Geneva	CH	1953-07-01
CSTO	Moscow	R	2002-10-07
EU	Brussels	B	1992-02-07
NATO	Brussels	B	1949-09-17
PCA	null	null	1899-07-29
WFTU	Prague	CZ	1945-10-03
⋮			

country

name	code	capital	area	population
Czech Republic	CZ	Prague	78,703	10,321,120
Switzerland	CH	Bern	41,290	7,207,060
Russia	R	Moscow	17,075,200	148,178,487
Belgium	B	Brussels	30,510	10,170,241
Turkey	TR	Ankara	780,580	62,484,478
United Kingdom	GB	London	244,820	58,489,975
Egypt	ET	Cairo	1,001,450	63,575,107
⋮				

is_member

country	organization	type
CZ	CERN	member
CZ	EU	member
CZ	WFTU	member
CH	CERN	member
R	CERN	observer
B	CERN	member
B	EU	member
B	NATO	member
TR	CERN	observer
TR	NATO	member
TR	WFTU	member
GB	C	member
GB	CERN	member
GB	EU	member
GB	NATO	member
⋮		

borders

country1	country2	length
GR	TR	206
CZ	A	362
CZ	D	646
CZ	PL	658
FL	CH	41
SK	CZ	215
CH	F	573
CH	A	164
CH	D	334
CH	I	740
PL	R	206
UA	R	1,576
B	F	620
B	D	167
B	NL	450
L	B	148
TR	IR	499
⋮		

is_member(organization) $\overset{fk}{\Rightarrow}$ organization(abbreviation)
is_member(country) $\overset{fk}{\Rightarrow}$ country(code)
borders(country1) $\overset{fk}{\Rightarrow}$ country(code)
borders(country2) $\overset{fk}{\Rightarrow}$ country(code)
organization(country) $\overset{fk}{\Rightarrow}$ country(code)

1 The following parts all refer to the **mondial** relational schema on Page 1.

a Write an RA query that returns the scheme (abbreviation, established, country) listing the abbreviation of organisations, when the organisation was established, and the code of the country which is a full member of the organisation.

b Write a query in each of the following query languages that returns the scheme (name), listing the name of countries that are not a full member of any organization.

i) RA

ii) Datalog

iii) SQL

c Write an SQL query returning the scheme (name, population, border_length) listing the name and population of every country in the database, together with the total length of any land borders of that country.

d Consider the following SQL query:

```
SELECT organization.abbreviation
FROM   organization
WHERE  NOT EXISTS (SELECT *
                   FROM   is_member
                          JOIN country
                          ON   is_member.country=country.code
                   WHERE  is_member.organization=organization.abbreviation
                   AND    country.area<=40000
                   AND    is_member.type='member')
```

i) Briefly explain the semantics of the query, and compute the result of the query on the fragment of data given on Page 1.

ii) Rewrite the query listed above to not include the NOT or EXCEPT operators.

e Write an SQL query returning the scheme (organization, name, population, pc_total) listing the name and population of each country that is full member of an organisation. The pc_total gives the percentage that the population of the country represents relative to the total population of all members of the organisation.

The five parts carry, respectively, 10%, 30%, 20%, 20%, and 20% of the marks.

2 a We wish to setup a new database containing information about members of a college, and examinations in the college.

Members of the college are identified by their cid number, and have their name and date of birth recorded. Members may be classified as either staff or student. For staff members we record their office room number, and the phone number.

Courses in the college have a unique code, a full title, and a term number in which they run. A course may be replaced by one other course, and a course may be the replacement for more than one course.

Each course may have any number of exam papers in a given year (and may have no exam papers), with each exam paper in a given year having a number and a maximum mark. Exam papers are not shared between courses, and we wish to record the maximum mark of each exam paper.

Students have their current degree programme recorded, and might also have the end date of their studies recorded. For each student we record one member of staff as being their personal tutor. We make a record of the mark achieved by each student in each exam paper.

i) Design an ER$^{\mathcal{ADHKLMNOSVW}}$ schema to represent this new database.

ii) Map the ER schema you designed in (i) into a relational schema.

b Suppose that a relation $R(A, B, C, D, E, F, G, H)$ has the functional dependencies:
$S = \{A \rightarrow CE, AE \rightarrow F, BD \rightarrow BG, C \rightarrow CAF, E \rightarrow F, G \rightarrow AD\}$.

Also suppose it has been proposed that the relation R be decomposed into $R_1(AG), R_2(ACEF), R_3(BDG)$

i) Compute a minimum cover S_c of S.

ii) Identify and justify all the candidate keys of R.

iii) Determine and justify if each of R_1, R_2, and R_3 is in 3NF. If the relations are not in $3NF$, give an alternative decomposition of R that is in 3NF.

c The following histories describe the sequence of operations performed respectively by three transactions T_1–T_3.

$H_1 = r_1[c_{CZ}], w_1[c_{CZ}], r_1[c_R], w_1[c_R], c_1$
$H_2 = r_2[c_{GB}], r_2[c_B], r_2[c_R], r_2[c_{CZ}], c_2$
$H_3 = r_3[c_R], w_3[c_R], r_3[c_B], w_3[c_B], c_3$

For each of the following concurrent executions, briefly explain if execution is serialisable and recoverable, and identify which (if any) anomalies occur.

i) $H_a = r_1[c_{CZ}], w_1[c_{CZ}], r_1[c_R], r_3[c_R],$
$w_3[c_R], r_3[c_B], w_3[c_B], w_1[c_R], c_3, c_1$

ii) $H_b = r_3[c_R], w_3[c_R], r_2[c_{GB}], r_2[c_B], r_2[c_R], r_2[c_{CZ}],$
$r_1[c_{CZ}], w_1[c_{CZ}], r_1[c_R], w_1[c_R], c_1, r_3[c_B], w_3[c_B], c_3, c_2$

The three parts carry, respectively, 40%, 40%, and 20% of the marks.