IMPERIAL COLLEGE LONDON

TIMED REMOTE ASSESSMENTS 2020-2021

BEng Honours Degree in Computing Part I
MEng Honours Degrees in Computing Part I
for Internal Students of the Imperial College of Science, Technology and Medicine

This paper is also taken for the relevant assessments for the
Associateship of the City and Guilds of London Institute

PAPER COMP40007

INTRODUCTION TO DATABASES

Wednesday 12 May 2021, 10:00
Duration: 95 minutes
Includes 15 minutes for access and submission

Answer ALL TWO questions
Open book assessment

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 types of membership include member for a full member, observer for non-members with a right to attend meetings, associate for partial members with some voting rights, *etc*.

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 $(abbreviation, established, name)$ listing the abbreviation of organisations, when the organisation was established, and the names of countries that are full members of that organization.

b   Write an SQL query that returns the scheme $(name, population)$, listing the name and population of island states: *i.e.* those countries with no land border.

c   Write an SQL query returning the scheme $(organization, population)$ listing organizations with at least 20 full members, together with the total population of those full members.

d   Consider the following SQL query:

```
SELECT    abbreviation
FROM      organization
WHERE     NOT EXISTS (SELECT type
                      FROM    is_member
                      EXCEPT
                      SELECT type
                      FROM    is_member
                      WHERE   is_member.organization=organization.abbreviation)
ORDER BY abbreviation
```

    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 in the RA.

    iii)   Rewrite the query listed above in Datalog.

e   Write an SQL query returning the scheme $(abbreviation, established, no\_member, no\_observer, no\_other)$ listing the abbreviation and establishment date of every organization, together with the number of countries that are members, that are observers, and that have any other type of membership.

The five parts carry, respectively, 10%, 15%, 20%, 35%, and 20% of the marks.

2 a   We wish to setup a new database containing information about placements at companies undertaken by students of a college as part of their degree programme.

Each student is identified by their cid number, and we record their current year of study. If the student has gradudated, then the year of graduation is recorded.

We record the various sites that a company has, identifying each site by the combination of the company name and the postcode of the site. We also record the town that the site is located in (or the nearest town if in the countryside).

Certain sites are regarded as approved sites for undertaking placements, and we record the start and end dates for which the approval holds.

We record all the placements that a student undertakes, for each placement we record the start date, duration and year of study during which the placement was undertaken. There are rules that no student may return to the same site for a second placement, and that placements can only take place at approved sites.

For some placements the student will have a member of college staff to act as a supervisor. Each member of staff is identified by their email address, and we record their name and position in the college.

   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 $R(A, B, C, D, E, F, G)$ has the functional dependencies:
$S = \{A \rightarrow BDE, AC \rightarrow F, BC \rightarrow FG, D \rightarrow A, G \rightarrow FG, GF \rightarrow C\}$.

Also suppose it has been proposed that the relation R be decomposed into $R_1(ABDE), R_2(BCFG)$

   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 and R_2 are in 3NF. If the relations are not in $3NF$, give an alternative decomposition of R that is in 3NF.

   iv) Determine and justify if each of R_1 and R_2 are in BCNF. If the relations are not in BCNF, give an alternative decomposition of R that is in BCNF.

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

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

Give a concurrent execution H_d of all three transactions that results in a deadlock state, and draw the waits-for graph (WFG) for that deadlock state.

The three parts carry, respectively, 35%, 45%, and 20% of the marks.