IMPERIAL COLLEGE OF SCIENCE, TECHNOLOGY AND MEDICINE

EXAMINATIONS 2012

BSc Honours Degree in Mathematics and Computer Science Part III
MSci Honours Degree in Mathematics and Computer Science Part III
MSc in Computing Science
for Internal Students of the Imperial College of Science, Technology and Medicine

*This paper is also taken for the relevant examinations for the
Associateship of the Royal College of Science*

PAPER C526

DATABASES

Friday 4 May 2012, 10:00
Duration: 120 minutes

Answer THREE questions

Paper contains 4 questions
Calculators not required

Several parts of the following questions make use of the **mondial** relational database, a fragment of which is listed below, containing some information about countries and the organizations those countries are members of. The encompasses table details the percentage of the area of a country that is within a continent. Note that a question mark after a column name indicates that the column is nullable, and any primary key columns of each table are underlined.

encompasses		
country	continent	percentage
AUS	Australia/Oceania	100.00
CH	Europe	100.00
DK	Europe	100.00
FL	Europe	100.00
IS	Europe	100.00
N	Europe	100.00
NZ	Australia/Oceania	100.00
S	Europe	100.00
SF	Europe	100.00
USA	America	100.00
ET	Africa	90.00
ET	Asia	10.00
TR	Asia	97.00
TR	Europe	3.00

organization		
abbreviation	city?	established?
ANZUS	Canberra	1951-09-01
EFTA	Geneva	1960-01-04
NATO	Brussels	1949-09-17
NC	Stockholm	1952-03-16

country				
name	code	capital	area	population
Australia	AUS	Canberra	7686850.00	18260863
Switzerland	CH	Bern	41290.00	7207060
Denmark	DK	Copenhagen	43070.00	5249632
Liechtenstein	FL	Vaduz	160.00	31122
Iceland	IS	Reykjavik	103000.00	270292
Norway	N	Oslo	324220.00	4383807
New Zealand	NZ	Wellington	268680.00	3547983
Sweden	S	Stockholm	449964.00	8900954
Finland	SF	Helsinki	337030.00	5105230
United States	USA	Washington	9372610.00	266476278
Turkey	TR	Ankara	780580.00	62484478
Egypt	ET	Cairo	1001450.00	63575107

is_member	
country	organization
AUS	ANZUS
CH	EFTA
DK	NATO
DK	NC
FL	EFTA
IS	EFTA
IS	NATO
IS	NC
N	EFTA
N	NATO
N	NC
NZ	ANZUS
S	NC
SF	NC
TR	NATO
USA	ANZUS
USA	NATO

located					
city	province	country	river?	lake?	sea?
Copenhagen	Denmark	DK	NULL	NULL	Baltic Sea
Helsinki	Uusimaa	SF	NULL	NULL	Baltic Sea
Oslo	Oslo	N	NULL	NULL	North Sea
Stockholm	Stockholm	S	NULL	Maelaren	Baltic Sea
Reykjavik	Iceland	IS	NULL	NULL	Atlantic Ocean

The following foreign key relationships exist between the tables:

is_member(organization) $\overset{fk}{\Rightarrow}$ organization(abbreviation)

is_member(country) $\overset{fk}{\Rightarrow}$ country(code)

encompasses(country) $\overset{fk}{\Rightarrow}$ country(code)

located(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 lists the names of countries, the organization they belong to, and the city in which that organization is based.

b   Write an RA query that lists the names of all cities that appear in located or in country.

c   Consider the following RA query:

is_member $\div$ $\pi_{organization}$ $\sigma_{country='DK'}$ is_member

   i)   List the result of the query, and explain the semantics of the query.

   ii) Translate the RA query into an equivalent SQL query.

   iii) Translate the RA query into an equivalent Datalog query.

d   Write a query in each of the following languages to find the capital cities of European countries (*i.e.* those countries with at least 50.0% of their area within Europe) which do not have an organization based in that city.

   i)   RA

   ii) SQL

   iii) Datalog

e   Give an RA query using just the primitive RA operators equivalent to the following SQL query.
```
SELECT capital
FROM   country
INTERSECT
SELECT city
FROM   organization
```

The five parts carry, respectively, 10%, 10%, 35%, 30%, and 15% of the marks.

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

a   Consider the following SQL query:

```
SELECT country.name
FROM    country
WHERE  country.code=SOME(SELECT is_member.country
                          FROM    is_member
                          JOIN is_member AS denmark
                          ON is_member.organization=denmark.organization
                          WHERE denmark.country='DK')
```

   i)   Compute the result of the query on the fragment of the mondial data given on Page 1, and briefly explain its semantics.

   ii) Rewrite the query into an equivalent query that does not use the SOME operator or any nested SELECT statements.

b   Write an SQL query to list the abbreviation of oraganizations that have or might have been established after 1959, and which contain European members (defined as countries with at least 50% of their area within Europe).

c   Write an SQL query that lists the names of continents, together with the total area of each continent computed by taking the sum of the percentage area of each country in each continent. The listing should be in ascending order of area.

d   i)   Write an SQL query that returns one row per organization, and lists the organization, the city where it is based, the number of European members, the number of African members, and the number of Asian members. (To qualify as a member belonging to a continent, a country must have at least 50.0% of its area within the continent).

   ii) What change would you make to your SQL query in order to make it only return rows where organizations have at least one European member?

e   Write an SQL query which returns rows with the name of a country, the capital city of the country, and the abbreviation of any organization based in that capital city (*i.e.* returning one row for each such organization). Should no organizations be based in a particular capital city, then the country name and capital city must be returned in one row, with the text 'No Organizations' in the organization column. The results should be ordered by country and organization.

The five parts carry, respectively, 25%, 15%, 15%, 30%, and 15% of the marks.

3 a   Suppose you have to design a new database to hold information about the registration of companies in countries, together with information of the contracts covering mergers between companies.

Each country is identified by a code, and for each country we need to store its name, and optionally its capital city. Each company is identified by its name, and we store the company registration number, total number of shares, and the country of registration of the company.

Certain companies are registered as banks, and for those companies we also store the captial and the tier 1 capital of the bank.

A merger between two (or more) companies will be governed by a contract. For each contract we give a serial number, and store the date of the contract, the country the contract is legally made in, and possibly an end date of the contract. At least one bank will advise on the merger, each bank charging a separate fee to the two (or more) companies involved in the merger.

   i)   Design an ER schema to represent this new database.

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

   b   The following histories describe the sequence of operations performed by three transactions.

$T_1 : r_1(c_{ch}), r_1(c_{is}), r_1(c_n), r_1(c_{fl})$

$T_2 : r_2(c_n), w_2(c_n), r_2(c_s), w_2(c_s), r_2(c_{ch}), w_2(c_{ch})$

$T_3 : w_3(c_{sf}), w_3(c_{is}), w_3(c_{fl}), w_3(c_{usa})$

   i)   Briefly explain if the following concurrent execution of T_1, T_3 is serialisable and recoverable.

   $w_3(c_{sf}), w_3(c_{is}), r_1(c_{ch}), r_1(c_{is}), w_3(c_{fl}), r_1(c_n), r_1(c_{fl}), c_1, w_3(c_{usa}), c_3$

   ii) Briefly explain if the following concurrent execution of T_1, T_3 is serialisable and recoverable.

   $w_3(c_{sf}), w_3(c_{is}), r_1(c_{ch}), r_1(c_{is}), r_1(c_n), r_1(c_{fl}), w_3(c_{fl}), w_3(c_{usa}), c_3, c_1$

   iii) Briefly explain if a concurrent execution of just T_1 and T_3 could ever deadlock.

   iv) Give a concurrent execution of T_1, T_2, T_3 which produces a deadlock involving all three transactions, and draw a waits-for graph for the deadlock situation.

The two parts carry equal marks.

4a Suppose that a relation $R(A, B, C, D, E, F, G, H)$ has the functional
dependencies:
$$S = \{AB \rightarrow ACDE, ABD \rightarrow CEF, GH \rightarrow FE, E \rightarrow F, D \rightarrow B,$$
$$C \rightarrow ABCD\}.$$

    i) Compute a minimum cover S_c of S.

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

    iii) Decompose the relation R into 3NF.

    iv) Decompose the relation R into BCNF.

    v) Identify which (if any) of the FDs in S_c are not preserved by the BCNF you
have decomposed from R.

b The table below lists the contents of a database log, which keeps only UNDO
records.

LOG	b_1
UNDO	$w_1[c_{fl}, \text{population}=31122]$
UNDO	$w_1[c_{sf}, \text{population}=5096672]$
LOG	b_2
UNDO	$w_2[c_{aus}, \text{population}=18260863]$
UNDO	$w_2[c_{usa}, \text{population}=266476278]$
UNDO	$w_2[c_{sf}, \text{population}=5105230]$
UNDO	$w_1[c_{aus}, \text{population}=19284997]$
LOG	c_2
LOG	b_3
UNDO	$w_1[c_{usa}, \text{population}=268467239]$
UNDO	$w_3[c_{usa}, \text{population}=268539881]$

    i) If the country table on disc was found to have the data listed as on Page 1,
describe the actions performed by the recovery procedure, and what
population figures will be left in country after recovery.

    ii) Considering the time when c_2 occurs, describe and justify what must have
been written to disc, which updates might have been written to disc, and
what must not have been written to disc.

The two parts carry, respectively, 70%, and 30% of the marks.