IMPERIAL COLLEGE OF SCIENCE, TECHNOLOGY AND MEDICINE

EXAMINATIONS 2013

BEng Honours Degree in Information Systems Engineering Part II
MEng Honours Degree in Information Systems Engineering Part II
BSc Honours Degree in Mathematics and Computer Science Part II
MSci Honours Degree in Mathematics and Computer Science Part II
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 City and Guilds of London Institute
This paper is also taken for the relevant examinations for the Associateship of the Royal College of Science

PAPER C526

DATABASES

Wednesday 1 May 2013, 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 **stock** relational database, a fragment of which is listed below. The database holds data about companies, which are classified as being in sectors such as retail, oil, *etc*. For public companies, the database records the stock exchanges on which shares in the company are traded, together with the ticker code used to identify that company on a particular exchange. The database also records which country a stock exchange is based (in the column based), and which country a company is headquartered (in the column hq). A record is also kept of offices a company holds in different countries.

company

cname	hq?	sector
ARM	GB	Tech
Apple	US	Tech
Arcadia	GB	Retail
BP	GB	Oil
Ford	US	Manu
HP	US	Tech
John Lewis	GB	Retail
Ryanair	IE	Air
Shell	NL	Oil
Tesco	GB	Retail

country

iso_code	short_name	trade_block?
GB	United Kingdom	EU
FR	France	EU
NL	Netherlands	EU
US	United States	null
IE	Ireland	EU

exchange

xname	based
AIM	GB
Dublin	IE
LSE	GB
NYSE	US
Nasdaq	US

office

cname	iso_code
HP	FR
HP	GB
HP	IE
HP	US
Ford	GB
Ford	US
John Lewis	IE
John Lewis	GB
ARM	GB
ARM	US
Ryanair	GB
Ryanair	IE
Ryanair	FR

public_company

cname	legal_status	net_assets
ARM	plc	738
Apple	inc	70532
BP	plc	48914
Ford	inc	9330
HP	inc	14023
Ryanair	plc	2713
Shell	plc	142744
Tesco	plc	14596

trades_on

xname	cname	ticker
Dublin	Ryanair	RY4B
LSE	ARM	ARM
LSE	BP	BP
LSE	Ryanair	RYA
LSE	Shell	RDSA
LSE	Tesco	TSCO
NYSE	BP	BP
NYSE	Ford	F
NYSE	HP	HPQ
Nasdaq	Apple	AAPL
Nasdaq	ARM	ARMH
Nasdaq	Ryanair	RYAAY

company(hq) $\stackrel{fk}{\Rightarrow}$ country(iso_code)
public_company(cname) $\stackrel{fk}{\Rightarrow}$ company(cname)
office(iso_code) $\stackrel{fk}{\Rightarrow}$ country(iso_code)
office(cname) $\stackrel{fk}{\Rightarrow}$ company(cname)
exchange(based) $\stackrel{fk}{\Rightarrow}$ country(iso_code)
trades_on(xname) $\stackrel{fk}{\Rightarrow}$ exchange(xname)
trades_on(cname) $\stackrel{fk}{\Rightarrow}$ public_company(cname)

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

a Write an RA query that returns the scheme (iso_code,cname,net_assets) listing the country of the stock exchange where a company is traded, together with the net assets of the company.

b Write an RA query that returns the scheme (cname,iso_code) listing company names and countries where they have offices, excluding the country where the company has its headquarters.

c Consider the following RA query:

$$\pi_{\text{cname}}\,\text{public\_company} - \pi_{\text{cname}}(\pi_{\text{cname,hq as iso\_code}}\,\text{company} \cap \pi_{\text{cname,based as iso\_code}}(\text{public\_company} \bowtie \text{trades\_on} \bowtie \text{exchange}))$$

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 that returns the scheme (cname,iso_code) listing companies that have offices in a country or have their shares traded on an exchange based in a country.

i) RA

ii) SQL

iii) Datalog

e Give an RA query, using appropriate primitive and derived RA operators, that is equivalent to the following SQL query.

```
SELECT public_company.cname
FROM   public_company
WHERE  NOT EXISTS (SELECT xname
                   FROM   trades_on
                   WHERE  trades_on.cname='BP'
                   EXCEPT
                   SELECT xname
                   FROM   trades_on
                   WHERE  trades_on.cname=public_company.cname)
```

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

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

a Consider the following SQL query:

```
SELECT company.hq AS iso_code,
       public_company.cname,
       public_company.net_assets
FROM   company
       JOIN public_company USING(cname)
WHERE  net_assets>=
           ALL (SELECT net_assets
                FROM   company AS same_country_company
                       JOIN public_company USING(cname)
                WHERE  same_country_company.hq=company.hq)
```

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 into an equivalent query that does not use the ALL or SOME operators.

b Write an SQL query that returns the scheme (cname,no_exchanges) listing the number of exchanges where the shares of a company are traded, listing only companies that are traded on more that one exchange.

c Write an SQL query that returns the scheme (iso_code,cname,pc) listing the country where public companies are headquartered, and the percentage of the total net assets of such companies in each country that a particular company's net assets represents.

d Write an SQL query that returns the scheme (xname,no_retail,no_oil) listing every exchange, together with the number of companies on that exchange that are in the retail or oil sectors, and which are also headquartered in country with ISO code GB.

e Write an SQL query that returns the scheme (cname) listing those companies that do not have an office in any country that might be in the EU trade block.

The five parts carry, respectively, 25%, 15%, 20%, 20%, and 20% of the marks.

3 a Suppose you have to design a new database to hold information about the countries of the world. The countries are divided into UN members and non-members. For all countries we identify them by their ISO code, and record an official name, and optionally a date the country became independent and a currency. We record which regions make up a country, identified within the country by a region name (but such names may also be used for regions in other countries). Each region has its population and area recorded.

For non-members, we record a comment about the country, and which UN members recognise the non-member, and the date that recognition was made.

For members, we record the date they joined the UN, the annual contribution made to the UN, and the membership of UN organisations. UN organisations are identified by their name, and we record the city where they are based.

Certain UN members are members of the security council, and we record the date the member joined the security council, and whether the member is permanent member of the security council.

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.

$H_1 = r_1[c_{US}], w_1[c_{US}], r_1[c_{GB}], w_1[c_{GB}], c_1$
$H_2 = r_2[c_{FR}], r_2[c_{US}], r_2[c_{GB}], r_2[c_{IE}], c_2$
$H_3 = r_3[c_{GB}], r_3[c_{IE}], r_3[c_{US}], w_3[c_{GB}], c_3$

i) Briefly explain if the following concurrent execution is serialisable and recoverable. If non-serialisable, explain what anomaly occurs.

$H_a = r_1[c_{US}], w_1[c_{US}], r_2[c_{FR}], r_2[c_{US}], r_2[c_{GB}], r_2[c_{IE}], c_2, r_1[c_{GB}], w_1[c_{GB}], c_1$

ii) Briefly explain if the following concurrent execution is serialisable and recoverable. If non-serialisable, explain what anomaly occurs.

$H_b = r_2[c_{FR}], r_2[c_{US}], r_3[c_{GB}], r_3[c_{IE}], r_2[c_{GB}], r_3[c_{US}], w_3[c_{GB}], c_3, r_2[c_{IE}], c_2$

iii) Briefly explain if the following concurrent execution is serialisable and recoverable. If non-serialisable, explain what anomaly occurs.

$H_c = r_1[c_{US}], w_1[c_{US}], r_1[c_{GB}], r_3[c_{GB}], r_3[c_{IE}], r_3[c_{US}], w_3[c_{GB}], w_1[c_{GB}], c_3, c_1$

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.

The two parts carry equal marks.

4a Suppose that a relation $R(A, B, C, D, E, F, G, H)$ has the functional dependencies:
$S = \{A \rightarrow DEGH, AH \rightarrow H, B \rightarrow F, CF \rightarrow AB, E \rightarrow G, EH \rightarrow D, GH \rightarrow A\}$.

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, maintaining FDs.

iv) Decompose the relation R into BCNF, and 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 of updates to the public_company table.

REDO	$w_1[p_{BP}, \text{net\_assets} = 45,000]$
REDO	$w_1[p_{HP}, \text{net\_assets} = 10,000]$
REDO	$w_2[p_{ARM}, \text{net\_assets} = 800]$
REDO	$w_2[p_{BP}, \text{net\_assets} = 46,000]$
REDO	$w_2[p_{HP}, \text{net\_assets} = 11,000]$
REDO	$w_1[p_{Ford}, \text{net\_assets} = 9,000]$
LOG	c_1
REDO	$w_2[p_{Ford}, \text{net\_assets} = 9,500]$
REDO	$w_3[p_{BP}, \text{net\_assets} = 47,000]$
LOG	c_3

i) If at the time of recovery the public_company table on disc was found to have the data listed as on Page 1, describe the actions performed by the recovery procedure, and what net_asset figures will be left after recovery.

ii) Considering the time just after when c_1 occurs, describe and justify which updates from the above log must have been written to disc, which might have been written to disc, and which must not have been written to disc.

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