IMPERIAL COLLEGE OF SCIENCE, TECHNOLOGY AND MEDICINE

EXAMINATIONS 2015

BEng Honours Degree in Electronic and Information Engineering Part II
MEng Honours Degree in Electronic and Information Engineering Part II
BEng Honours Degree in Mathematics and Computer Science Part II
MEng Honours Degree in Mathematics and Computer Science Part II
BEng Honours Degree in Mathematics and Computer Science Part III
MEng 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

PAPER C526

DATABASES

Wednesday 29 April 2015, 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 **airroute** relational database, a fragment of which is listed below. It contains in airline information about airlines, and in airport information about airports. The hq_in column of airline describes where the airline is headquartered. The serves relation records which routes are served by airlines between airports, with each pair of airports being recorded only once for any given airline. Thus the flight table recording actual flights on the routes served has a column dir recording if the flight is inbound or outbound on the route stored in serves.

airline

acode	pcode	start_year	no_aircraft	hq_in	aname
AF	CDG	1974	245	FR	Air France
BA	LHR	1974	297	GB	British Airways
KQ	NBO	1977	56	KE	Kenya Airways
BE	EXE	1979	70	GB	FlyBe

country

iso_code	name	flag_carrier?	population?
KE	Kenya	KQ	45010056
FR	France	AF	63929000
UG	Uganda	null	35873253
GB	United Kingdom	BA	64100000

airport

pcode	elevation	iso_code	cname
EBB	3782	UG	Kampala
EXE	102	GB	Exeter
NBO	5327	KE	Nairobi
LGW	203	GB	London
LHR	83	GB	London
CDG	392	FR	Paris

serves

acode	pcode1	pcode2
BA	LHR	NBO
BA	EBB	LHR
BA	LHR	CDG
AF	LHR	CDG
KQ	LHR	NBO
KQ	EBB	NBO
KQ	NBO	CDG
BE	LGW	EXE

city

iso_code	cname	is_capital
KE	Nairobi	t
GB	London	t
GB	Newquay	f
GB	Exeter	f
FR	Paris	t
UG	Kampala	t

flight

fcode	acode	pcode1	pcode2	dir
KQ0101	KQ	LHR	NBO	O
KQ0102	KQ	LHR	NBO	I
KQ0112	KQ	NBO	CDG	O
BA0065	BA	LHR	NBO	O
BA0064	BA	LHR	NBO	I
BA0304	BA	LHR	CDG	O
BA0306	BA	LHR	CDG	O
BA0308	BA	LHR	CDG	O
BA0303	BA	LHR	CDG	I
BA0307	BA	LHR	CDG	I
BA0309	BA	LHR	CDG	I
BE0120	BE	LGW	EXE	I
BE0129	BE	LGW	EXE	O

flight_day

fcode	day
KQ0112	Fri
KQ0112	Sat
KQ0112	Sun
KQ0112	Tue
KQ0112	Thr

main_hub

pcode	icao_code
NBO	HKJK
LHR	EGLL
CDG	LFPG

airline(pcode) $\stackrel{fk}{\Rightarrow}$ main_hub(pcode)
airline(hq_in) $\stackrel{fk}{\Rightarrow}$ country(iso_code)
country(flag_carrier) $\stackrel{fk}{\Rightarrow}$ airline(acode)
airport(iso_code,cname) $\stackrel{fk}{\Rightarrow}$ city(iso_code,cname)
serves(acode) $\stackrel{fk}{\Rightarrow}$ airline(acode)
serves(pcode1) $\stackrel{fk}{\Rightarrow}$ airport(pcode)
serves(pcode2) $\stackrel{fk}{\Rightarrow}$ airport(pcode)
city(iso_code) $\stackrel{fk}{\Rightarrow}$ country(iso_code)
flight(acode,pcode1,pcode2) $\stackrel{fk}{\Rightarrow}$ serves(acode,pcode1,pcode2)
flight_day(fcode) $\stackrel{fk}{\Rightarrow}$ flight(fcode)
main_hub(pcode) $\stackrel{fk}{\Rightarrow}$ airport(pcode)

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

a Write an RA query that returns the scheme (aname,name,icao_code) listing the names of airlines established after 1970, together with the name of the country they are headquartered in, and the icao_code of the airline's main hub.

b Write an RA query that returns the scheme (name) listing the name of those countries in which no airlines are headquarted.

c Consider the following RA query:

$$\pi_{\text{acode}}\,\text{airline} - \pi_{\text{acode}}\,\sigma_{\text{airport}_a.\text{pcode}=\text{serves.pcode1} \wedge \text{airport}_b.\text{pcode}=\text{serves.pcode2} \wedge \text{airport}_a.\text{iso\_code}=\text{airport}_b.\text{iso\_code}}(\text{airport}_a \times \text{serves} \times \text{airport}_b)$$

i) List the result of the query when run on the fragment of data on Page 1, 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 returning the scheme (acode,pcode1,pcode2) listing each airline route in serves for which there is not at least one inbound and one outbound flight.

i) RA

ii) SQL

iii) Datalog

e Suppose the RA query airline $\bowtie$ airport has been executed at some point in time, after which Δ_a has been added to airline to give airline′, and Δ_p has been added to airport to give airport′. Give an RA query in terms of airline, airport, Δ_a and Δ_p that returns the additional rows returned by the query airline′ $\bowtie$ airport′.

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

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

a Consider the following SQL query:

```
SELECT acode
FROM   airline
EXCEPT
SELECT flag_carrier
FROM   country
```

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) Briefly explain if changing the query to use EXCEPT ALL could ever lead to different answers for any dataset held in the schema on Page 1.

iii) Rewrite the query listed above to use NOT IN instead of the EXCEPT operator.

iv) Rewrite the query listed above to use NOT EXISTS instead of the EXCEPT operator.

b Write an SQL query that returns the scheme (name,low,high) listing every country in the database, together with the number of low airports and the number of high airports in that country. A low airport has an elevation of less than 1,000 feet, and a high airport has an elevation of 1,000 feet or more.

c Write an SQL query that returns the scheme (aname,name,pc_country,pc_world) listing each airline, and the country in which the airline is headquartered. The pc_country column is the number of aircraft of the airline as a percentage of the total aircraft headquarted in the country, and the pc_world column is the number of aircraft of the airline as a percentage of the total aircraft in the world.

d Write an SQL query that returns the scheme (acode,aname) listing airlines that serve every capital city recorded in the database.

e Consider the following SQL query:

```
SELECT name,
       COUNT(country.iso_code) AS no_entities,
       SUM(no_aircraft) AS no_aircraft
FROM   country
       LEFT JOIN (SELECT hq_in AS iso_code,
                         no_aircraft
                  FROM   airline
                  UNION ALL
                  SELECT iso_code,
                         0
                  FROM   airport) AS aviation_entity
                  ON   country.iso_code=aviation_entity.iso_code
GROUP BY name
ORDER BY name
```

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 UNION operator.

The five parts carry, respectively, 20%, 15%, 15%, 20%, and 30% of the marks.

3 a Suppose you have to design a new database to hold the following information about the sun, planets, dwarf planets, asteroids, and moons of the solar system.

Each such body in the solar system is given a unique name, has known mass, and mean diameter. Most bodies also have a year of discovery. Apart from the sun, all bodies orbit some other body, with some mean distance between the two bodies, and a period of the orbit.

Some pairs of bodies form a binary-system, where they not only each orbit some third body, but also orbit each other.

A number of space probes have been sent into the solar system, each identified by a mission name, with known launch dates, and for some probes, a date of last contact. Each probe has various modules, each module being given a description, and numbered consecutively on each probe. We know the power consumption of each module.

Each probe will target one or more body in the solar system, with the closest distance of the probe to the body recorded, together with the date of that distance being achieved. For images made by probes, we wish to generate a serial number for the image (unique over all images), and record the target the image was for, the date the image was made, and the URL from which the image may be retrieved.

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 four transactions $T_1 - T_4$.

$H_1 = r_1[c_{KE}], w_1[c_{KE}], r_1[c_{FR}], w_1[c_{FR}], c_1$
$H_2 = r_2[c_{KE}], r_2[c_{GB}], r_2[c_{FR}], c_2$
$H_3 = r_3[c_{GB}], r_3[c_{FR}], r_3[c_{UG}], c_3$
$H_4 = r_4[c_{UG}], w_4[c_{UG}], r_4[c_{FR}], r_4[c_{GB}], r_4[c_{KE}], w_4[c_{KE}], c_4$

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

$$H_a = r_4[c_{UG}], w_4[c_{UG}], r_4[c_{FR}], r_4[c_{GB}], r_4[c_{KE}], r_1[c_{KE}], w_1[c_{KE}], r_1[c_{FR}], w_1[c_{FR}], c_1, w_4[c_{KE}], c_4$$

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

$$H_b = r_3[c_{GB}], r_4[c_{UG}], w_4[c_{UG}], r_3[c_{FR}], r_3[c_{UG}], c_3, r_4[c_{FR}], r_4[c_{GB}], r_4[c_{KE}], w_4[c_{KE}], c_4$$

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

$H_c = r_1[c_{KE}], w_1[c_{KE}], r_2[c_{KE}], r_1[c_{FR}], r_2[c_{GB}], r_2[c_{FR}], c_2, w_1[c_{FR}], c_1,$

iv) Briefly explain which pair of the transactions taken from $T_1 - T_4$ will be serialisable and recoverable for all concurrent executions of the pair.

v) Give a concurrent execution of the four transactions, which produces a deadlock involving all four transactions, and draw a waits-for graph for the deadlock state.

The two parts carry, respectively, 45%, and 55% of the marks.

4a Suppose that a relation $R(A, B, C, D, E, F, G)$ has the functional dependencies:
$S = \{A \rightarrow BCH, C \rightarrow H, D \rightarrow E, DG \rightarrow A, E \rightarrow F,$
$EFG \rightarrow ADE, F \rightarrow E\}$.

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 records of updates to the airline table, and where CP is a cache consistent check point record.

UNDO	w_3[$\mathsf{a_{AF}}$, no_aircraft = 242]
REDO	w_3[$\mathsf{a_{AF}}$, no_aircraft = 240]
UNDO	w_4[$\mathsf{a_{BE}}$, no_aircraft = 67]
REDO	w_4[$\mathsf{a_{BE}}$, no_aircraft = 70]
UNDO	w_2[$\mathsf{a_{BA}}$, no_aircraft = 295]
REDO	w_2[$\mathsf{a_{BA}}$, no_aircraft = 297]
UNDO	w_1[$\mathsf{a_{AF}}$, no_aircraft = 240]
REDO	w_1[$\mathsf{a_{AF}}$, no_aircraft = 245]
CP	{1,2,3}
UNDO	w_3[$\mathsf{a_{BA}}$, no_aircraft = 297]
REDO	w_3[$\mathsf{a_{BA}}$, no_aircraft = 299]
UNDO	w_1[$\mathsf{a_{KQ}}$, no_aircraft = 55]
REDO	w_1[$\mathsf{a_{KQ}}$, no_aircraft = 56]
LOG	c_1
UNDO	w_2[$\mathsf{a_{KQ}}$, no_aircraft = 56]
REDO	w_2[$\mathsf{a_{KQ}}$, no_aircraft = 57]
LOG	c_2

i) Just after the time of the checkpoint CP, what updates *must* be on disc, and what updates *might* be on disc.

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

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