IMPERIAL COLLEGE OF SCIENCE, TECHNOLOGY AND MEDICINE

EXAMINATIONS 2017

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

Thursday 4 May 2017, 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 **property** relational database used by an Estate Agent, a fragment of which is listed below. It contains information about **property**, including the area in which each property is located, the **agent** who is responsible for marketing a property, the **rates** (property tax) payable per year for the property and, if it is a rental property, the maintenance charge (in **maint**) per year for the property.

Companies that take on a lease of a property may have a schedule of rent payments recorded, stating the date on which rent is due. If a company is owned by another company in the database, then we record it as being a subsidiary of that other company.

area

outer_pc	aname	county
E8	East Ham	London
SW7	South Kensington	London
SW19	Wimbledon	London
WC1V	Holborn	London
KT18	Epsom and Ewell	Surrey
KT19	Epsom and Ewell	Surrey

property

land_reg	rates	agent	outer_pc	inner_pc	pname?
SGL120	23000.00	JJS	SW19	2PA	Acme House
SGX149	33000.00	FB	SW19	2PL	Morris House
TGW992	120000.00	JJS	WC1V	1PQ	Bush House
TPA458	30000.00	JJS	SW7	4JS	null
UVB877	3000.00	FB	KT19	4JS	null
RXL224	7500.00	FB	KT18	5AA	null

company

cno	cname	subsidiary_of?
12345	Big Inc	null
84587	Acme Computing	12345
95050	Tiny Inc	84587
48067	Bloggs Ltd	null
56567	IC	null

lease

cno	land_reg	start_date	end_date
12345	TGW992	2016-01-01	2025-12-31
95050	TGW992	2010-05-01	2015-09-30
12345	SGX149	2014-01-01	2018-12-31
56567	TPA458	2010-05-01	2018-12-31

rent

rno	cno	land_reg	due	amount
1	12345	TGW992	2016-03-25	50000.00
2	12345	TGW992	2016-06-24	50000.00
1	95050	TGW992	2010-06-24	35000.00
2	95050	TGW992	2010-09-29	35000.00
3	95050	TGW992	2010-12-25	35000.00
4	95050	TGW992	2011-03-25	35000.00
5	95050	TGW992	2011-06-24	35000.00

rental

land_reg	maint
SGL120	3500.00
SGX149	6000.00
TGW992	19200.00
RXL224	4500.00
TPA458	4500.00

telephone

cno	number
12345	020 8542 1234
12345	020 8542 1400
12345	020 8542 1401
84587	020 7222 1234
84587	020 7222 1200
95050	020 7224 9000

lease(cno) $\overset{fk}{\Rightarrow}$ company(cno)

lease(land_reg) $\overset{fk}{\Rightarrow}$ rental(land_reg)

rent(cno,land_reg) $\overset{fk}{\Rightarrow}$ lease(cno,land_reg)

rental(land_reg) $\overset{fk}{\Rightarrow}$ property(land_reg)

property(outer_pc) $\overset{fk}{\Rightarrow}$ area(outer_pc)

company(subsidiary_of) $\overset{fk}{\Rightarrow}$ company(cno)

telephone(cno) $\overset{fk}{\Rightarrow}$ company(cno)

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

  a   Write an RA query that returns the scheme (cname,due,amount) listing the names of companies, and rent payments due from the company for all properties with an outer post code of WC1V.

  b   Write a query in each of the following languages returning the scheme (agent,land_reg,pname,maint) listing those rental properties that have not had any rent due since the 1st January 2014.

      i)   RA

      ii) SQL

      iii) Datalog

  c   Write a query in each of the following languages returning the scheme (agent) listing the agents that manage properties in every county recorded in the database.

      i)   RA

      ii) SQL

      iii) Datalog

  d   Consider the following RA query:
$$\pi_{cname}(company \bowtie lease \bowtie (\pi_{land\_reg} \text{ rent} -$$
$$(\pi_{land\_reg} \text{ rent} - \pi_{land\_reg} \sigma_{country='London'}(area \bowtie rental))))$$

      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.

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

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

a   Consider the following SQL query:

```
SELECT  company.cno ,
        start_date ,
        rental.land_reg
FROM    company
        NATURAL LEFT JOIN lease
        NATURAL FULL OUTER JOIN rental
```

   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 an equivalent query that does not use
        OUTER JOIN.

b   Write an SQL query that returns the scheme (cno, cname, no_of_phones) that
     lists every company with at least two telephone numbers, where no_of_phones is
     the number of such telephone numbers.

c   Write an SQL query that returns the scheme (cno, cname) listing every company
     that has no subsidiaries recorded in the database.

d   Write an SQL query that returns the scheme (agent, no_london, no_surrey) listing
     agents who had at least one of their properties with a lease active on the 1st
     January 2017, and giving the number of such properties in London and in Surrey.

e   Write an SQL query that returns the scheme
     (land_reg, maint, cno, pc_company, county, pc_county) listing every property
     with a lease active on the 1st January 2017, together with the company it is
     leased to, and the county the property is within. The pc_company gives the
     percentage of all maintenance charges for the company (at that date) that the
     maintenance of this property represents. Similarly, pc_county gives the
     percentage of all maintenance charges (at that date) for the county that the
     maintenance represents.

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

3 a   Suppose you have to design a new database for a share registrar, which needs to hold the following information about the companies with shares traded on public stock exchanges.

The share registrar has a number of customers, which may be divided into individuals and companies. We identify each customer by their hid, and record their name, address, and legal country of residence. For individuals we record their national insurance number, and for companies their company registration number.

Some companies are public companies with shares listed on one or more stock exchange, and we must record the capital of such companies. When listed on an exchange the company is given a ticker code to use on that exchange, and we record how many shares of the company are listed on that exchange. For each customer we record the percentage of a listed company's total shares that are held by that customer.

Each stock exchange is identified by a code, and we record the name of the exchange, and the city and country where the exchange is based. Listed companies may issue at most one action per day for their shares listed on a particular stock exchange. For such actions, we record the date and type of the action. If the action has a deadline, then this is also recorded.

Countries have the currency, name and ISO code recorded.

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

   ii) Suppose you are required to map the ER schema you designed in (i) into a relational schema. List that part of the schema that holds details of the stock exchanges, companies listed on those exchanges, the corporate actions on those listings, and any relationships between those three things. You listing must include all foreign keys that involve any tables that you list, even if one of the tables in the foreign key is not listed.

   b   Suppose that a relation $R(A, B, C, D, E, F, G)$ has the functional dependencies $S = \{AC \to BEF, ACD \to DE, B \to AD, E \to BDA, F \to FG\}$.

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

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

4a   The following histories describe the sequence of operations performed respectively by three transactions T_1–T_3.

$$H_1 = r_1[c_{L1}], w_1[c_{L1}], r_1[c_{W7}], w_1[c_{W7}], r_1[c_{E8}], w_1[c_{E8}], c_1$$
$$H_2 = r_2[c_{W7}], r_2[c_{SW7}], r_2[c_{E8}], c_2$$
$$H_3 = r_3[c_{W7}], r_3[c_{E8}], w_3[c_{E8}], r_3[c_{SW7}], w_3[c_{SW7}], c_3$$

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

$$H_a = r_3[c_{W7}], r_3[c_{E8}], r_1[c_{L1}], w_1[c_{L1}], r_1[c_{W7}], w_1[c_{W7}], r_1[c_{E8}],$$
$$w_3[c_{E8}], r_3[c_{SW7}], w_3[c_{SW7}], c_3, w_1[c_{E8}], c_1$$

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

$$H_b = r_2[c_{W7}], r_2[c_{SW7}], r_1[c_{L1}], w_1[c_{L1}], r_1[c_{W7}], w_1[c_{W7}],$$
$$r_1[c_{E8}], w_1[c_{E8}], r_2[c_{E8}], c_2, c_1$$

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

$$H_c = r_2[c_{W7}], r_1[c_{L1}], w_1[c_{L1}], r_1[c_{W7}], w_1[c_{W7}], r_2[c_{SW7}],$$
$$r_1[c_{E8}], r_2[c_{E8}], c_2, w_1[c_{E8}], c_1$$

  iv) If the following history were attempted to be executed using Strong Strict 2PL, explain which (if any) operation would be blocked. If any operation is blocked, write down the actual order of execution of operations that would be performed by the database, and state what the serialisation order of the two transactions is.

$$H_d = r_1[c_{L1}], w_1[c_{L1}], r_2[c_{W7}], r_2[c_{SW7}], r_1[c_{W7}], w_1[c_{W7}],$$
$$r_1[c_{E8}], r_2[c_{E8}], w_1[c_{E8}], c_1, c_2$$

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

 b   The table below lists the contents of a database log, which uses a cache consistent checkpoint procedure, and which records updates to the rental table on Page 1.

```
LOG     b1
UNDO    w1[rSGL120, maint = 3500.00]
REDO    w1[rSGL120, maint = 3600.00]      LOG     checkpoint(1,2)
LOG     b2                                 UNDO    w2[rRXL224, maint = 4400.00]
UNDO    w2[rTPA458, maint = 3100.00]       REDO    w2[rRXL224, maint = 4500.00]
REDO    w2[rTPA458, maint = 3300.00]       LOG     c2
LOG     b3                                 UNDO    w1[rSGX149, maint = 6000.00]
UNDO    w3[rSGL120, maint = 3600.00]       REDO    w1[rSGX149, maint = 6450.00]
REDO    w3[rSGL120, maint = 3700.00]       LOG     b4
UNDO    w1[rTGW992, maint = 19200.00]      UNDO    w4[rTPA458, maint = 3300.00]
REDO    w1[rTGW992, maint = 20500.00]      REDO    w4[rTPA458, maint = 4600.00]
                                           LOG     c4
                ⋮
```

i)   If the rental table on disc was found to have the data listed as on Page 1,
     describe the actions performed by the recovery procedure, and give the
     recovered version of the rental table.

ii) If no REDO log entries are made, what additional actions are taken by the
     database to maintain recoverability, and how will the recovery procedure of
     (i) differ.

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