UNIVERSITY OF PORTSMOUTH

BACHELOR OF SCIENCE (HONS) IN COMPUTING

AUGUST 2004 EXAMINATION

17 AUGUST 2004
IT354 DATABASE DESIGN & MANAGEMENT
MARKING SCHEME

TIME: 3 HOURS

Students are permitted to refer the following books and course material:

1. "Fundamentals of Database systems ", Elmasri and Navathe, Addison Wesley.
2. "Database Systems – A practical Approach to Design, implementation and management", Thomas Connolly and Carolyn Begg, Addison Wesley
3. Course Contents.

NOTE: No handwritten notes and printed materials other than the above are permitted.

☞ **INSTRUCTIONS TO CANDIDATES:**

- Choose any **THREE (3)** questions out of **FIVE (5)** questions.
- Each question is worth 33 marks. 1 extra mark will be added to all candidates scripts to construct a percentage.
- **DO NOT** answer more than **THREE (3)** questions in total.
- Please start every question on a new page.
- Answers will not be marked if they are illegible.
- Write the question numbers (in the order you have attempted them) in the boxes provided in the answer script.

Write your **INDEX NUMBER** and **MODULE NUMBER** on the cover page of the answer booklet(s).

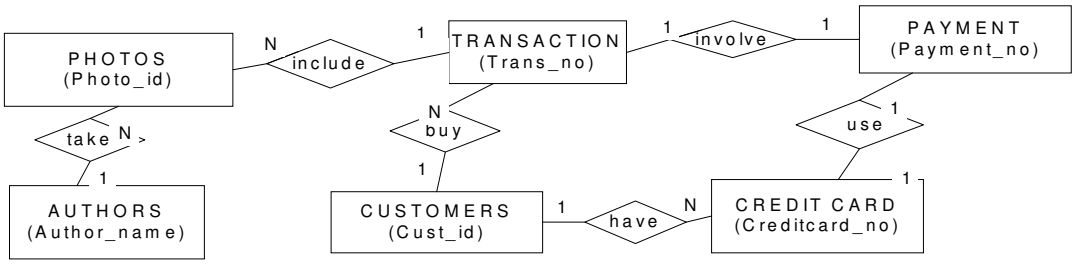
QUESTION 1

a) An Art gallery maintains a database of photographs that can be bought over the internet.

The database has three functions

i) maintaining the catalogue with entities *photograph* and *photographer*
ii) selling the photos with entities *transaction* and *customer*
iii) maintaining the payment particulars with entities *payment* and *credit card details*

The ER diagram of the existing system is shown in the diagram below. The primary keys are given in parenthesis.



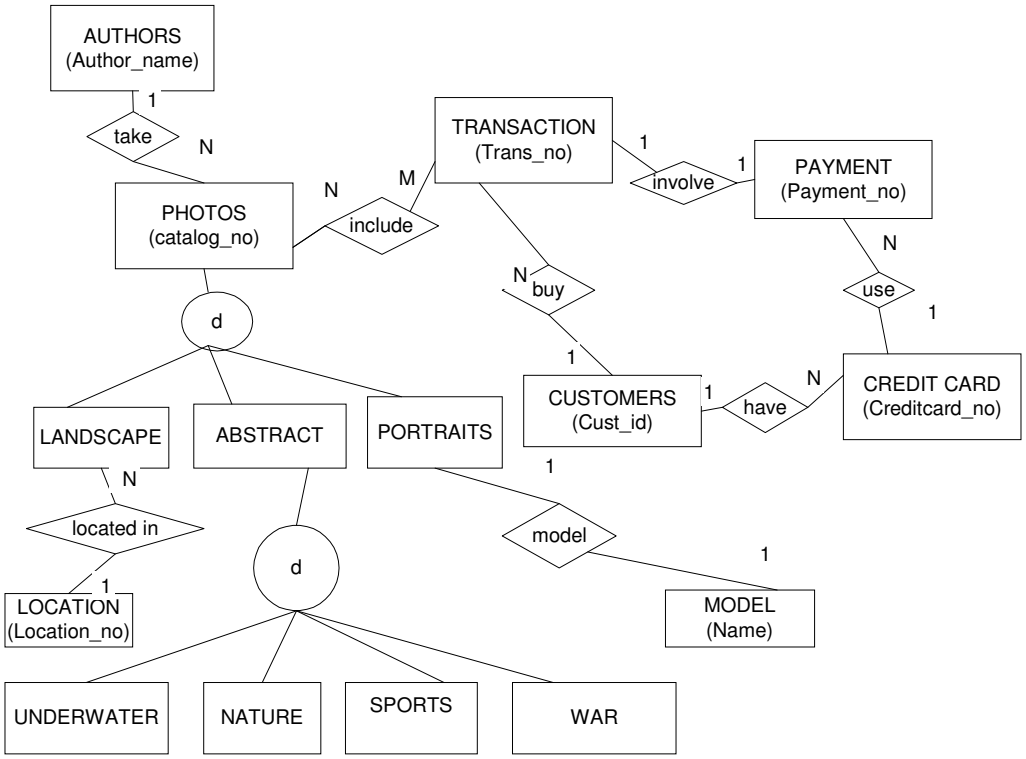
You are required to enhance this ER model to an EER one by adding the following details:-

1) The database will store some technical information for every photo, like the type (colour or black and white), type of film, shutter speed, f-stop, and price.

2) The photographs are to be grouped into three categories: Landscapes, Portraits and General (everything else). Landscape has information about the location of photo where it was taken. Portrait has information about the model used The General category can be easily extended to describe other types of photographs such as Underwater, Nature, Sports, War etc.

A location is identified by a unique number and a model by their name.

a) EER model



5 marks for each of the 2 super/sub types and 5 marks for adding in the Location and Model entities.
Total 15 marks

b) Consider the following two relations *R* and *S*:

A	B
a	b
c	b
d	e

R

B	C
b	c
e	a
b	d

S

Compute:

(i) $R \bowtie S$

(ii) $S \div \{c, d\}$

(iii) $S \div \{b, c\}$

(iv) $R \bowtie_{B<C} S$ (where < means alphabetic order on letters)

(i)

A	B	C
a	b	c
a	b	d
c	b	c
c	b	d
d	e	a

6 marks

(ii) b

4 marks

(iii) Empty set

4 marks

(iv)

A	B	C
a	b	c
a	b	d
c	b	c
c	b	d

4 marks

[TOTAL FOR QUESTION 1: 33 MARKS]

QUESTION 2

a) A small database with only three relations, R1, R2 and R3 is distributed over three sites, S1, S2 and S3. You are to allocate each of the three relations R1, R2 and R3 to one of the three sites S1, S2 and S3 using the nonredundant best fit method, (which is one technique which helps DBAs to decide how to distribute data across sites).

There are only three transactions run on this database: T1, T2 and T3.
The originating sources for each transaction (that is the sites at which they are run) and the percentage frequency with which the transactions run are shown below:-

Transaction sources and frequencies

T1	sources S1 and S2	20%
T2	sources S1 and S3	30%
T3	sources S2 and S3	50%

Note that transactions are equally likely to originate from any of their sources.

The usage made of each relation by each transaction is given by:-

Transaction behaviour

T1 2 reads to R1 and 2 writes toR1
2 reads to R2

T2 3 reads to R1
2 reads to R3 and 1 write to R3

T3 2 reads to R2 and 1 write to R2
2 reads to R3

		S1	S2	S3
R1	T1	4x 20/2	4x 20/2	----
	T2	3x 30/2	-----	3x 30/2
	T3	-----	-----	-----
R2	T1	2x 20/2	2x 20/2	-----
	T2	-----	----	-----
	T3	------	3x 50/2	3x 50/2
R3	T1	-----	-------	-----
	T2	3x 30/2	------	3x 30/2
	T3	-------	2x 50/2	2x 50/2

Which gives:

	S1	S2	S3
R1	40 + 45 =85	40	45
R2	20	20 +75 = 95	75
R3	45	50	45 + 50=95

Hence

R1 at S1, R2 at S2, R3 at S3

Up to 5 marks for working it out
10 marks for correct formula, but wrong application/errors
15 marks for correct solution as above.

b) Consider the following set, *F*, of functional dependencies:

$AB \rightarrow C$ $\quad$ $D \rightarrow EG$
$C \rightarrow A$ $\quad$ $BE \rightarrow C$
$BC \rightarrow D$ $\quad$ $CG \rightarrow BD$
$ACD \rightarrow B$ $\quad$ $CE \rightarrow AG$

Two analysts, Mike and Dave have each produced a different minimal cover for *F* (given below).
Decide and give reasons for your decision whether each or either of these actually is a minimal cover for *F*.

Mike's "minimal cover":
AB → C
C → A
BC → D
CD → B
D → E
D →G
BE → C
CG → D
CE →G

Dave's "minimal cover":
AB → C
C → A
BC → D
D → E
D →G
BE → C
CG → B
CE →G

We can use IR4(Decomposition) to split RHSs and get:-

AB → *C*	*D* → *E*	*CG* → *D*
C → *A*	*D* → *G*	*CE* → *A*
BC → *D*	*BE* → *C*	*CE* → *G*
ACD → *B*	*CG* → *B*	

Clearly *CE* → *A* is redundant as it is implied by *C* → *A*.
CG → *B* is redundant since *CG* →*D, C* → *A* and *ACD* → *B* imply *CG* → *B*
No more dependencies are redundant.

However *ACD* → *B* can be replaced by *CD* → *B*, since *C* → *A* and so *CD* → *B* can be deduced from *ACD* → *B* and *C* → *A*
No further reduction is possible.
This is Mike's set.

OR
Eliminating *CE* → *A*, (as before), *CG* → *D* (because *BC* → *D and CG* → *B)* and *ACD* → *B* (because *C* → *A* and *D* → *G* and *CG* → *B*) from *F***,** gives us Dave's minimal set.

Hence both sets are minimal.

9 marks for correctly establishing that each set is a minimal set.
Total 18 marks.

[TOTAL FOR QUESTION 3: 33 MARKS]

QUESTION 3

Tomes On Line, an internet only bookseller, requires a database to keep information on all of its offered books, together with their authors, publishers, and warehouse locations where the books are stocked.

For each publisher, the database must record the publisher code, name, and the publisher's address(es). A publisher can have several different addresses. No two publishers have the same name.

For each book, the database must record its ISBN code, title, the publisher code, the price, and published format (Possible formats are: hardback, paperback and/or audioCD). Note that for any title published in several formats, each separate format has a distinct ISBN code. In addition, the database must record the identity number and name of each of the authors of the book. If there is more than one author, they must be listed in the order in which they are listed on the book. This may or may not be ordered alphabetically. Furthermore the warehouse location and number in stock at that warehouse are also recorded for each book.

For each warehouse, the database must record the identity number, the name, the location, and the number of employees at that warehouse.

a) Functional dependencies

i) PUBLISHER_CODE → PUBLISHER_NAME, ADDRESS
ii) PUBLISHER_NAME → PUBLISHER_CODE, ADDRESS
iii) WAREHOUSE_NUMBER → NAME, LOCATION, NUMBER_EMPLOYEES)
iv) ISBN_CODE → TITLE, PUBLISHER_CODE, FORMAT, PRICE
v) AUTHOR_NUMBER→ AUTHOR_NAME
vi) ISBN_CODE, AUTHOR_NUMBER → SEQ_NUMBER
vii) WAREHOUSE_NUMBER, ISBN_CODE → UNITS_IN_STOCK

2 marks for each correct FD, up to a maximum of 14 marks

a) BCNF relations

PUBLISHER (***PUBLISHER_CODE***, PUBLISHER_NAME, ADDRESS)

WAREHOUSE (***WAREHOUSE_NUMBER***, NAME, LOCATION, NUMBER_EMPLOYEES)

BOOK (***ISBN_CODE***, TITLE, PUBLISHER_CODE, FORMAT, PRICE)

AUTHOR (***AUTHOR_NUMBER***, AUTHOR_NAME)

WROTE (***ISBN_CODE, AUTHOR_NUMBER***, SEQ_NUMBER)

STOCK (***WAREHOUSE_NUMBER, ISBN_CODE***, UNITS_IN_STOCK)

1 mark for each relation and each PK, total 6 marks

b) A previous attempt at constructing this database had produced the relation

WROTE (ISBN_CODE, AUTHOR_NUMBER, FORMAT) with some sample data as shown in the table below:-

ISBN_CODE	AUTHOR_NUMBER	FORMAT
0-321-44567	3345	Hardback
0-321-44567	3345	Paperback
0-321-44567	2178	Hardback
0-321-44567	2178	Paperback
0-321-44567	4356	Hardback
0-321-44567	4356	Paperback
0-456-38921	3345	Hardback

Show that a multi-valued dependency exists within this relation and normalize it to 4NF.

ISBN_CODE multidetermines AUTHOR_NUMBER and FORMAT, because the values of AUTHOR_NUMBER and FORMAT determined by ISBN_CODE are independent of each other, that is to say that if, say row 2 were missing from the table we could reconstruct it from the other rows present.

5 marks for a complete and correct explanation of the MVD. Incomplete answers to be marked on their merits.

The 4NF for this relation is:-

ISBN_CODE	AUTHOR_NUMBER
0-321-44567	3345
0-321-44567	3345
0-321-44567	2178
0-321-44567	2178
0-321-44567	4356
0-321-44567	4356
0-456-38921	3345

ISBN_CODE	FORMAT
0-321-44567	Hardback
0-321-44567	Paperback
0-321-44567	Hardback
0-321-44567	Paperback
0-321-44567	Hardback
0-321-44567	Paperback
0-456-38921	Hardback

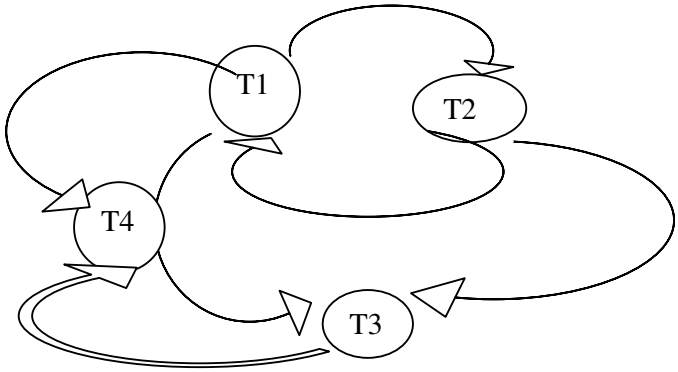
4 marks for each relation, total 8 marks.

[TOTAL FOR QUESTION 3: 33 MARKS]

QUESTION 4

a) Draw the conflict graph for the following schedule, S, of four transactions:
S: W1(x)W2(x)R3(x)R3(y)R1(y)R1(x)W4(y)
Hence determine whether or not the schedule is serializable

Schedule S is not serialisable



7 marks for the graph and 3 marks for "not serializable"

b) Consider the following schedule of four transactions. Show how this schedule would be executed under strict 2PL (two phase locking).

R1(y)W4(x) R4(y)R3(w)R2(y)W1(y)W1(x)W2(x)W2(z)W3(x)R4(z)

T1	**T2**	**T3**	**T4**
Write-lock (y)			
R(y)			
WAIT		Read-lock (w)	
WAIT	WAIT	R(w)	
W(y)	WAIT	WAIT	
Write-lock(x)	WAIT	WAIT	
W(x)	WAIT	WAIT	
UNLOCK(y)	WAIT	WAIT	
UNLOCK(x)	Read-lock(y)	WAIT	read-lock (y)
	R(y)	WAIT	Write-lock (x)
	WAIT	WAIT	W(x)
	WAIT	WAIT	R(y)
	WAIT	WAIT	Read-lock(z)
	WAIT	WAIT	R(z)
	WAIT	WAIT	UNLOCK(z)
	WAIT	WAIT	UNLOCK(y)
	WAIT	WAIT	UNLOCK(x)
	WAIT	Write-lock(x)	
	WAIT	W(x)	
	WAIT	UNLOCK(x)	
	Write-lock (x)	UNLOCK(w)	
	W(x)		
	Write-lock(z)		
	W(z)		
	UNLOCK(z)		
	UNLOCK(x)		
	UNLOCK(y)		

3 marks for each transaction correctly shown, total 12 marks

c) i) Show that when a strict 2PL is applied to the following schedule, C, deadlock occurs.

C = R1(x)R2(y)W2(y)R2(x)W2(x)R1(y)

T1	T2
Read-lock(x)	
R(x)	Write-lock(y)
	R(y)
	W(y)
	Read-lock(x)
	R(x)
	WAIT
Wait	WAIT
WAIT	Wait

T2 waits for x to be unlocked by T1, and T1 waits for y to be unlocked by T2

2 marks for T1, 2 marks for T2, and 1 mark for showing how the deadlock occurs.
Total 5 marks.

ii) Consider the following protocol:
"If T_j is not blocked (i.e. not waiting for some locked item), then T_i is blocked and allowed to wait; otherwise abort T_i and restart T_i after a time delay" (the cautious waiting protocol).
What happens if we apply this protocol to the schedule C?

T1	T2
Read-lock(x)	
R(x)	Write-lock(y)
BLOCKED & WAITING	R(y)
BLOCKED & WAITING	W(y)
BLOCKED & WAITING	Read-lock(x)
BLOCKED & WAITING	R(x)
BLOCKED & WAITING	ABORT
Read-lock(y)	
R(y)	
UNLOCK (x)	
UNLOCK (y)	
	RESTART
	Write-lock(y)
	R(y)
	W(y)
	Write-lock(x)
	R(x)
	W(x)
	UNLOCK(x)
	UNLOCK(y)

3 marks for each transaction, total 6 marks

[TOTAL FOR QUESTION 4: 33 MARKS]

QUESTION 5

a) Student should identify any two strategies, calculate the costs and identify one of the strategies as costing less.

Example:
Strategy 1: Transfer all the relations and records in Laotan and Elharib to Smale and execute the query.

Transfer all the relation in Laotan to Smale
Cost of data transfer: (100 x 60 + 2000 x 30) x 1/1000= $66

Transfer all the relation in Elharib to Smale
Cost of date transfer = (20 x 30 + 1000 X 60) x 1/1000 = $60.60

Total cost = $66+ $60.6 = $126.60

Strategy 2:
Laotan: Project Driver_id attribute alone from Driver and transfer to Smale
Transfer the relation Driver_Vehicle relation to Smale
Coat of data transfer = (100x 15 + 2000 X 30) X 1/1000= $61.50

Elharib: Select one record from Vehicle with Vechicle_id = SH4000
Join the result with relation Trips.
Select the records corresponding to date condition.
Max number of records can be 90 (90 days).
Project three attributes vehicle_id, Driver_id and Trip_no

Cost of data transfer = (90 X 30) x 1/1000 = $2.70
Total cost = 61.50+ 2.70= $64.20

So strategy 2 is optimum.

2 marks for the description of each strategy, total 4 marks
3 marks for each cost calculation, total 6 marks
1 mark for the correct optimum answer.

Total 11 marks

If the student shows that any optimum query must follow the pattern
i) SELECT and PROJECT first (where ever possible) and then
ii) JOIN operation
Then 5 marks may be given irrespective of the cost calculation

b) i) Identify the predicate used to create the following horizontal fragments

Predicate used is : **Course_id** **1 mark**

ii) Write the relational algebra expressions to create the horizontal fragments with **Course_id** as predicate

$F1 = \sigma_{Course_id = IT354}$ (Student_course)
$F2 = \sigma_{Course_id = IT355}$ (Student_course)
$F3 = \sigma_{Course_id = IT356}$ (Student_course)
$F4 = \sigma_{Course_id = IT352}$ (Student_course)
$F5 = \sigma_{Course_id = IT358}$ (Student_course) **1 mark**

iii) Show whether the given fragments satisfy complete, disjoint and reconstruction conditions.

1) The fragment F4 is not a horizontal fragmentation. Attribute date of birth is missing. **1 mark**

2) Completeness: Fragments are not complete. Fragment corresponding to the relational algebra expression F5 = $\sigma_{Course_id = IT38}$ (Student_course) is missing. **2 marks**

3) Disjoint ness: Fragments are not disjoint.
In Fragment F1 the record is irrelevant

030091015390	Joseph Robert	10-03-1978	IT355

2 marks

4) Reconstruction: Union of all the four fragments F1, F2 , F3 and F4 does not give the global relation Student_course due to the followings

- Missing record –

010191021234	Chai Wei Mng	26-10-1980	IS358

- Duplicated record

030091015390	Joseph Robert	10-03-1978	IT355

2 marks

c) i) List the names of doctors and patients in pairs, so that, for each pair, the patient is over 60 years old and sees the doctor in 1999.

R1 = $\sigma_{Age=60}$ (Patients)
R2 = $\sigma_{Date >= '01\text{-}01\text{-}1999' \text{ and } <= date'31\text{-}12\text{-}1999'}$ (Records)
R3= R2 $_{Records.Dr\ id = Doctors.Dr\ id}$ * Doctors
R4 = R1 $_{R1.Patient\ id = R3.\ Patient\ id}$ * R3
R5 = $\Pi_{Patients.Name,\ Doctors.Name}$ (R4)

1 mark for each correct line up to a maximum of 5 marks

ii) Find the names of those male patients who have never seen a doctor.

S1 = $\Pi_{patient_id}$ ($\sigma_{Sex = 'M'}$ (Patients))
S2 = S1 $_{R1 \cdot patient_id = Patients \cdot patient_id}$ * Patients
S3 = S2 - $\Pi_{patient_id}$ (Records)
S4 = $\Pi_{Patients.Name}$ S3

1 mark for each correct line up to a maximum of 4 marks

iii) Find the patient name and diagnosis for all patients treated by doctors with the specialty "Neuro surgery"

T1 = $\sigma_{Specialty = Neuro\ surgery}$ (Doctors)

$T2 = T1_{\text{Records.Dr id = Doctors.Dr id}} * \text{Records}$
$T3 = \pi_{\text{Patient id, Diagnosis}}\ T2$
$T4 = \pi_{\text{Name, Diagnosis}}\ (T3 * \text{PATIERNTS})$

1 mark for each correct line up to a maximum of 4 marks

[TOTAL FOR QUESTION 5: 33 MARKS]

- End of Paper -