Surname: _____

First Name: _____

Student Number: _____

Signature: _____

The University of Queensland

Mid-Semester Examination, 20, April 2004

INFS7907

Advanced Database Systems

Time: 1 hour 30 minutes for working
10 minutes perusal before exam begins

ALL questions to be answered
Questions carry the number of marks indicated
Total marks: 100

Answers are to be written in the space provided
Additional booklets are provided for rough working

Marker's use only

1	2	3	4	5	6	7	8	Total

Question 1 (5 marks)

(a) Is a Relational Database System catalog itself a database? If so, who maintains it?

Yes. It is a relational database which stores meta-data about all databases implemented/populated on this DBMS, information about applications, users, users right and privileges etc. Depending on the product, the system catalog contains over 30 tables.

It is maintained by the DBMS automatically by any user (DBA) request to modify either the existing tables schema, creating new tables, deleting existing tables or any other information entry or change.

(b) Give an example of an insert to the catalog.

a. **eg. 'Create/Drop table' statement**
b. **NOT insert/delete/update tuple statements**

Question 2 (10 marks)

A table *r*(A, B, C, D), with the primary key underlined, is fragmented into the following fragments:

$r_1 = \Pi_{AC}\, \sigma_{A<5}(r)$

$r_2 = \Pi_{AB}\, \sigma_{A>5}(r)$

$r_3 = \Pi_{ACD}(r)$

(a) Is the above fragmentation correct? Justify your answer.

If not, suggest a modification to the fragment definitions to get a correct fragmentation.

No.

Not reconstructible, complete, or disjoint.

Not reconstructible and complete because values of B with A <=5 are not included.

Not Disjoint as r1 and r3 cover values of C with A<5

(b) Reconstruct the original table *r* from the correct fragments in (a) using relational algebra expression.

Change to cover all possible values and ensure disjointness (eg. Change $r_1 = \Pi_{AB}\, \sigma_{A<=5}(r)$)

Question 3 (10 marks)

Determine the derived horizontal fragments for the relation DRIVER, given the following simple predicates:

P1: Price > $27,000

P2: Price < $7,200

P3: Price ≥ $7,200 AND Price ≤ $27,000

VEHICLE

Vehicle#	Model	Make	Price	Colour
310HTR	1987	Corona	7200	Blue
467IUD	1991	Corolla	14980	White
901JST	1998	Camry	32000	Red
012LOP	1997	Commodore	27889	White
001NRE	1972	Corolla	3500	Brown
190PLO	2000	Falcon	78600	Grey
738MNQ	1997	Camry	19005	Green
657VDR	1989	Accord	6900	Black
435GTY	1996	Barina	21900	Blue

DRIVER

Name	Address	Phone	Gender	Age	Annual Salary	Vehicle
John	35 Down St	32097653	M	35	20300	310HTR
James	67 High St	73659807	M	38	45000	467IUD
Kelly	190 Main Road	73749101	F	46	59900	901JST
Fiona	4 Small St	23464839	F	33	67890	012LOP
Elvin	90 River View	98474733	M	21	1200	001NRE
Kalem	12 Sunset Ave	35875098	M	56	100900	190PLO
Emily	8 Staff House	76875443	F	29	36000	738MNQ
Annisa	76 Sixth Ave	91763549	F	27	19860	657VDR
Thomas	12 Warren St	25458346	M	36	39800	435GTY

P1:

Fiona	4 Small St	23464839	F	33	67890	012LOP
Kalem	12 Sunset Ave	35875098	M	56	100900	190PLO
Kelly	190 Main Road	73749101	F	46	59900	901JST

P2:

Elvin	90 River View	98474733	M	21	1200	001NRE
Annisa	76 Sixth Ave	91763549	F	27	19860	657VDR

P3:

John	35 Down St	32097653	M	35	20300	310HTR
James	67 High St	73659807	M	38	45000	467IUD
Emily	8 Staff House	76875443	F	29	36000	738MNQ
Thomas	12 Warren St	25458346	M	36	39800	435GTY

Question 4 (20 marks)

The following are four tables from a simplified university database:

`a = staff(name, department_name)`

`b = student(student_number, name, degree)`

`c = subject(code, name, number_of_credit, lecturer)`

`d = study(student, subject, result)`

where primary keys are underlined,

`study.student` **is a foreign key to** `student.student_number`**,**

`study.subject` **is a foreign key to** `subject.code`**,**

`subject.lecturer` **is a foreign key to** `staff.name`**.**

The database is fragmented according the following rules:

`staff` **is partitioned by three department "CSEE", "Maths", "Arts";**

`student` **is partitioned by three degree "MIT", "BIT", "BSc";**

`subject` **is partitioned by department which the lecturer works for;**

`study` **is not partitioned.**

(You may introduce your own names for the fragments).

For the following query

```
select  b.name
from    staff a, student b,  subject c, study d
where  b.degree = "MIT" and d.student = b.student_number and
          a.department = "Maths" and d.subject = c.code and
          c.lecturer = a.name
```

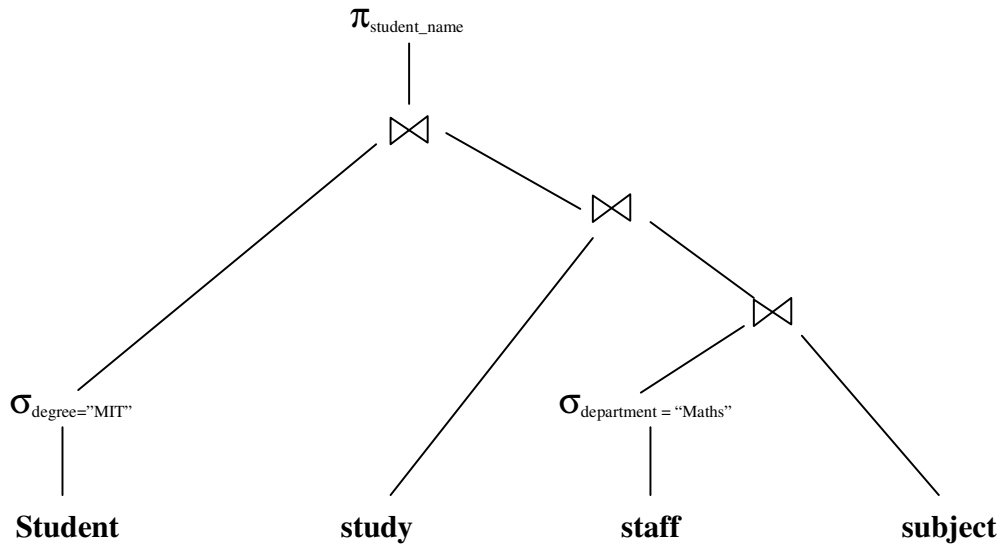
Give operator trees for the above query:

(i) after query decomposition (i.e., on global relations),

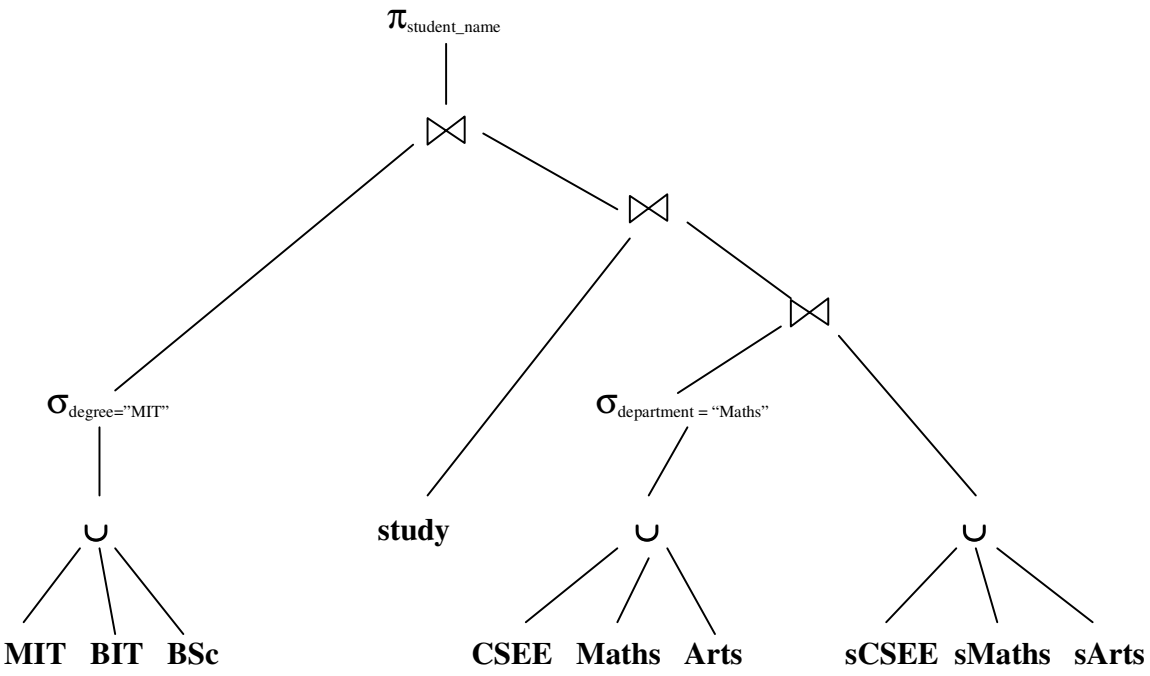
(ii) after the localization process,

(iii) after the reduction process.

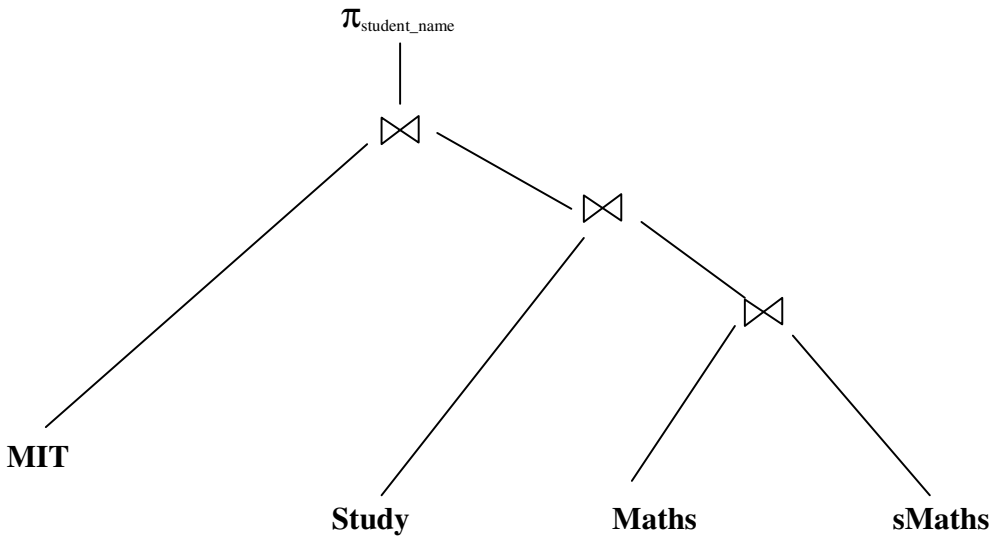
(i) after query decomposition (i.e., on global relations),



(ii) after the localization process,



(iii) after the reduction process.



Question 5 (8 marks)

In general, is the following equality true? Justify your answer.

$r \ltimes s = s \ltimes r$

In general this statement is not true. LHS "reduces' table r but the RHS table s.

It is sufficient to show any two tables with different schema to judge the difference.

Show by example:

r(A B C)	s(C D E)	$r \ltimes s =$	(A B C)
1 0 3	3 4 9		1 0 3
3 4 9	3 7 1		
6 2 1	2 1 3		

$s \ltimes r =$ (C D E)

3 4 9

3 7 1

Therefore in general $r \ltimes s \neq s \ltimes r$

Question 6 (20 marks)

(a) List the main differences in data consistency management in centralised and distributed Transaction Processing.

• Database Consistency
– No semantic integrity constraints are violated

• Transaction Consistency
– No incorrect updates during concurrent database accesses

• Mutual Consistency
– No difference in the values of a data item in all the replicas to which it belongs

Therefor have to ensure all updates are transmitted to all relevant sites

(b) Suppose that 2PC with Presumed Abort is used as the commit protocol. Explain how the system recovers from failure and deals with a transaction T when the coordinator site for T fails after writing a *commit* log record but before sending any further messages to its subordinates.

All the subordinates are blocked since they must have all voted yes for the coordinator to have decided to commit. Once the coordinator comes back up, it will find the commit log record, will redo T, and periodically resend a commit message to all its subordinates, until it receives an ack from all of them.

Question 7 (20 marks)

Assume that we need to build a new integrated system from two pre-existing databases; one managed and owned by Company A and another one by Company B, both operating in three Australian states only – QLD, ACT and NT.

Both companies store and maintain semantically compatible data although design in a different ways. The relational data structure for Company A and B is presented below. The data is a snapshot of tables for a given year (period of time)

Company A database – all sale values are in thousands (for example 23 stands for 23000);

Sale (ProductID, Sale-QLD , Sale-ACT, Sale-NT)

123	23	45	126
378	12	12	67
562	10	34	50
234	24	26	27
612	56	11	32
134	67	34	32

Company B database:

Sale (ProdID,	State,	Sale)
612	QLD	78000
612	NT	12000
234	ACT	34000
235	ACT	31000
134	QLD	10000
378	NT	9000
897	QLD	67000
562	ACT	78000

(a) Show a mapping of the data from Company A to data format of Company B using SQL and views.

Create view CompanyA-Sales (ProdID, State, Sale) as

(Select ProductID, 'QLD' , Sale-QLD*1000 from Sale@CompanyA)

UNION

(Select ProductID, 'ACT' , Sale-ACT*1000 from Sale@CompanyA)

UNION

(Select ProductID, 'NT' , Sale-NT*1000 from Sale@CompanyA);

(b) When the physical data is combined (from Company A database with Company B database) will the schema Sale (ProdID, State, Sale) be adequate for the integrated data store. If YES say why, if NO show further data modification required.

NO, because the sales for each state will have to be summed up to reflect the combined TOTAL sales for each state from the two companies.

Question 8 (7 marks)

Why in general do Federated Database Systems prove to give a better solution for system integration (by comparison to Multidatabse Systems)?

- **Easier maintenance (FD systems cooperate with other sites to ensure consistency) whereas MD Systems is difficult since global system has no control over local sites).**
- **Differences in data representation are resolved individually (eg each site can interpret export schema to it's own needs, so there is no need for global system develop a translation which is usable by all sites)**
- **No global queries. Global system does not have to deal with queries to the global system since all queries are local.**

The end of exam