



UNIVERSITY OF COLOMBO, SRI LANKA



UNIVERSITY OF COLOMBO SCHOOL OF COMPUTING



DEGREE OF BACHELOR OF INFORMATION TECHNOLOGY (EXTERNAL)

Academic Year 2006/2007 – 3rd Year Examination – Semester 6

IT6402 - Advanced Database Management Systems
Structured Question Paper

26th August, 2007
(THREE HOURS)

To be completed by the candidate

BIT Examination Index No:

Important Instructions:

- The duration of the paper is **3 (three) hours**.
- The medium of instruction and questions is English.
- This paper has **4 questions** and **16 pages**.
- **Answer all questions** (25 marks each).
- **Write your answers** in English using the space provided **in this question paper**.
- Do not tear off any part of this answer book.
- Under no circumstances may this book, used or unused, be removed from the Examination Hall by a candidate.
- Note that questions appear on both sides of the paper.
If a page is not printed, please inform the supervisor immediately.
- **Non-programmable Calculators may be used.**

Questions Answered

Indicate by a cross (X), (e.g. X) the numbers of the **four** questions answered.

	Question numbers			
	1	2	3	4
<u>To be completed by the candidate by marking a cross (X).</u>				
To be completed by the examiners:				

- 1) (a) Describe the following five terms numbered (i) to (v) used in concurrent execution of transactions.
(i) Serial Schedule

(02 marks)

ANSWER IN THIS BOX

Serial Schedule consists of a sequence of instructions from various transactions, where instructions belonging to a particular transaction appear always together in that schedule.

i.e.

Instructions of each transaction is executed consecutively without interleaving.

- (ii) Non-Serial Schedule

(02 marks)

ANSWER IN THIS BOX

Non-Serial Schedule consists of a sequence of instructions from various transactions, where all instructions belonging to some transactions do not appear together in that schedule.

i.e.

Instructions of each transaction are executed consecutively with some interleaving.

- (iii) Conflict-Serializable

(03 marks)

ANSWER IN THIS BOX

A schedule is said to be conflict-serializable when the schedule is serializable as well as it is conflict-equivalent to one or more serial schedules.

e.g. $\langle T1, T2 \rangle \sim \langle T2, T1 \rangle$

(iv) Unrecoverable Schedule

(03 marks)

ANSWER IN THIS BOX

If a transaction T1 aborts and a transaction T2 commits, but T2 relies on T1,
we have an unrecoverable schedule.

i.e.

A transaction committed transaction finds that the value read and commit had been
based on a dirty read.

(v) Cascading Aborts (Rollbacks) or Cascadeless

(02 marks)

ANSWER IN THIS BOX

A single transaction abort leads to a series of transaction rollbacks.

i.e.

Restrict schedules to read committed values to avoid cascadeless.

(b) Consider the following three transactions T1-T3 which are to be executed concurrently.

T1	T2	T3
READ (A) A:=A+20 WRITE (A) COMMIT	READ (A) A:=A+10 WRITE (A) READ (B) B:=B-10 WRITE (B) COMMIT	READ (A) READ (B) A:=A+B B:=0 WRITE (A) WRITE (B) ROLLBACK

The following six schedules S1-S6 are six different possibilities of executing two of the above transactions concurrently.

S1		S2	
T1: READ (A) A:=A+20 WRITE (A) COMMIT	T2: READ (A) A:=A+10 WRITE (A) READ (B) B:=B-10 WRITE (B) COMMIT	T1: READ (A) A:=A+20 WRITE (A) COMMIT	T2: READ (A) A:=A+10 WRITE (A) READ (B) B:=B-10 WRITE (B) COMMIT

S3		S4	
T1: READ (A) A:=A+20 WRITE (A) COMMIT	T3: READ (A) READ (B) A:=A+B B:=0 WRITE (A) WRITE (B) ROLLBACK	T1: READ (A) A:=A+20 WRITE (A) COMMIT	T2: READ (A) A:=A+10 WRITE (A) READ (B) B:=B-10 WRITE (B) COMMIT

S5		S6	
T1: READ (A) A:=A+20 WRITE (A) COMMIT	T2: READ (A) A:=A+10 WRITE (A) READ (B) B:=B-10 WRITE (B) COMMIT	T3: READ (A) READ (B) A:=A+B B:=0 WRITE (A) WRITE (B) ROLLBACK	T1: READ (A) A:=A+20 WRITE (A) COMMIT

(i) List the schedule(s) which illustrate(s) the problem of “lost update”.

(02 marks)

ANSWER IN THIS BOX

S1

(ii) List the schedule(s) which illustrate(s) the problem of “dirty read”.

(02 marks)

ANSWER IN THIS BOX

S3

(iii) List the non-serial schedules which are serializable.

(02 marks)

ANSWER IN THIS BOX

S4, S5, S6

(iv) List the conflict-serializable schedules.

(03 marks)

ANSWER IN THIS BOX

S2, S4, S5

(v) List the recoverable schedules.

(02 marks)

ANSWER IN THIS BOX

S2

(vi) List the Cascadeless schedules.

(02 marks)

ANSWER IN THIS BOX

S2, S6

- 2) (a) Cost based and heuristics based are two techniques to optimise a query. Briefly describe the two techniques.

(04 marks)

ANSWER IN THIS BOX

Cost-based approach searches for all possible plans and chooses the plan with the lowest cost.

Heuristics approach applies a set of rules to select the best plan.

- (b) A professional body conducts a competition annually under different categories. The SQL commands are used to create the database tables for this application.

```
CREATE TABLE Participants(pno VARCHAR(5), name VARCHAR(20),
address VARCHAR(30), telephone VARCHAR(10), email VARCHAR(15));
```

```
CREATE TABLE Categories(code VARCHAR(5), category_name
VARCHAR(15), description VARCHAR(30));
```

```
CREATE TABLE Winners(pno VARCHAR(5), code VARCHAR(5), year
INTEGER, award VARCHAR(6));
```

The following SQL-query is written to retrieve all Gold award winners with each winner's name, category and year.

```
SELECT name, category_name, year
FROM Participants p, Categories c, Winners w,
WHERE p.pno=w.pno and c.code=w.code and award="Gold"
```

Assume that there are 1000 participants, 20 categories and 10 Gold award winners out of 100 winners.

- (i) Express the above query using relational algebra. First apply more restrictive operations and finally eliminate the unwanted attributes.

(03 marks)

ANSWER IN THIS BOX

Gold_Winners = $\sigma_{award = 'Gold'}$ (Winners)

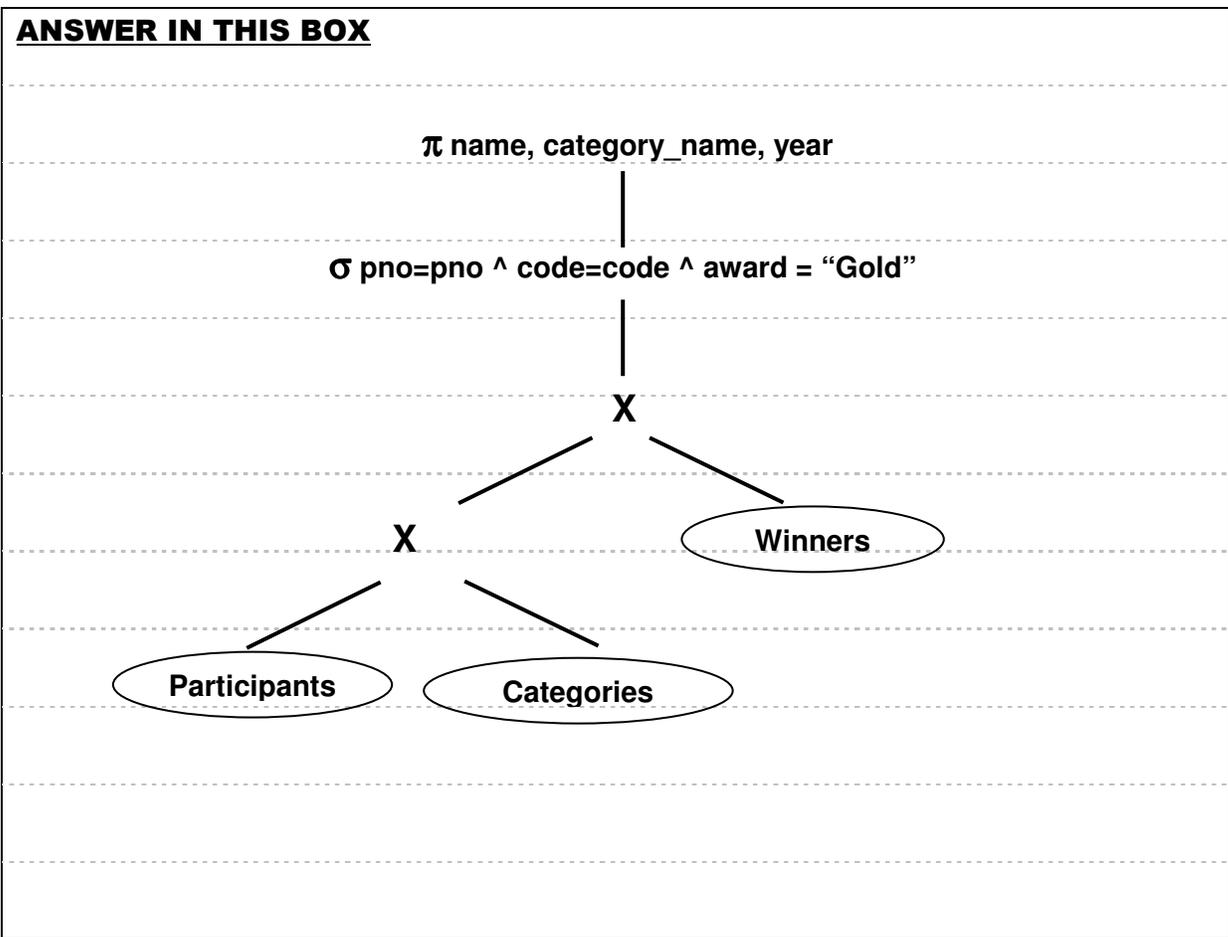
Gold_Win_Part = Participants $\bowtie_{pno = pno}$ Gold_Winners

Gold_Win_Det = Categories $\bowtie_{code = code}$ Gold_Win_Part

Result = $\pi_{name, category_name, year}$ (Gold_Win_Det)

(ii) Draw an initial query tree (in canonical form) for the above SQL-query.

(03 marks)



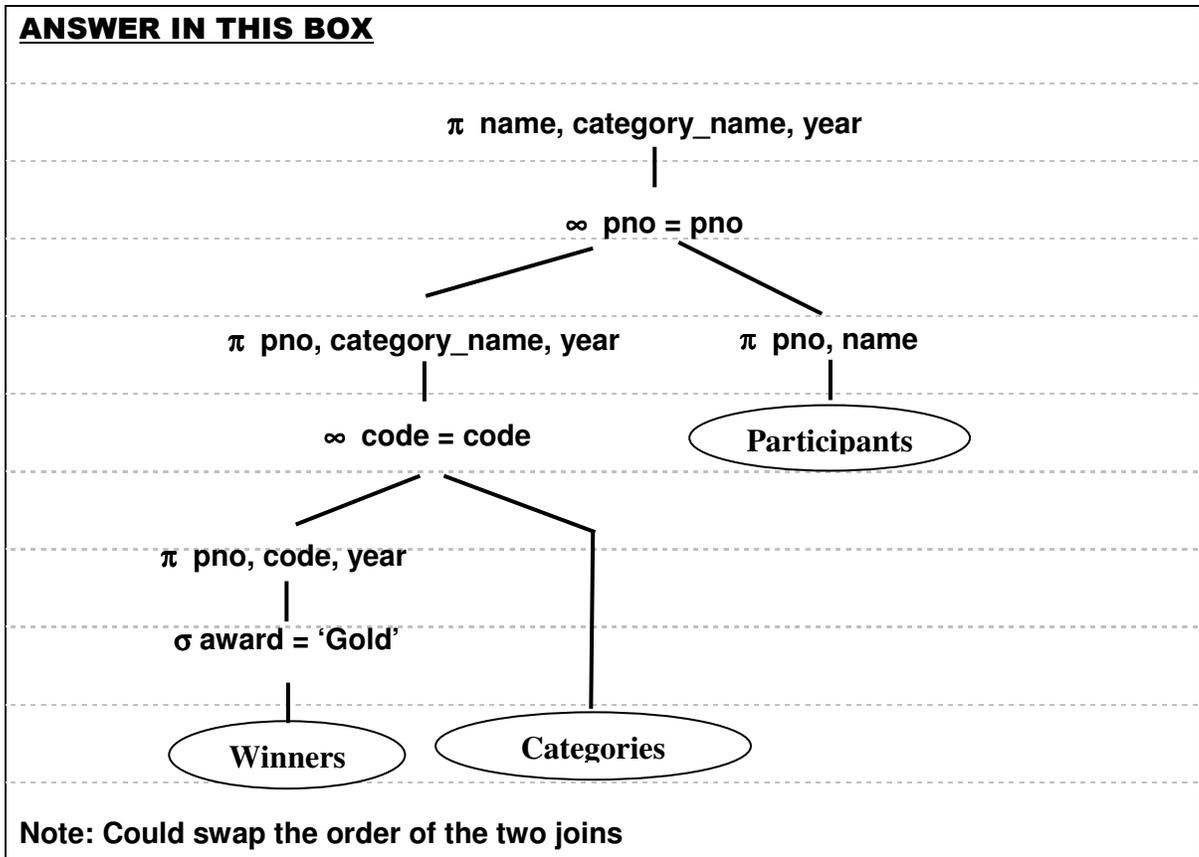
(iii) Apply heuristic optimisation to optimise the above query. Indicate the heuristics one would have to use to optimise the given query.

(04 marks)

- ANSWER IN THIS BOX**
- Moving SELECT operations down the query tree. i.e. award="Gold"
 - Applying the more restrictive SELECT operation first. i.e. Winners
 - Replacing CARTESIAN PRODUCT and SELECT with JOIN operations.
i.e. join by pno and code
 - Moving PROJECT operations down the query tree.

(iv) Draw the optimised query tree for the above query. Do not show the intermediate trees.

(04 marks)



(v) Derive in bytes the cost of processing each stage of the canonical query tree of part (ii).

(04 marks)

ANSWER IN THIS BOX

Record Length of Participants, Categories and Winners are 80, 50 and 20 respectively.

Cost of processing Participants = $80 * 1000 = 80,000$ bytes = 80kb

Cost of processing Categories = $50 * 20 = 1,000$ bytes = 1kb

Cost of processing Winners = $20 * 100 = 2,000$ bytes = 2kb [1]

Cross product = $(80+50+20)*1000*20*100=300,000,000=300Mb$ [1]

Selection = $150*10= 1,500$ bytes [1]

Projection = $39*10 = 390$ bytes [1]

(vi) Derive in bytes the cost of processing each stage of the optimised query tree of part (iv).

(03 marks)

<u>ANSWER IN THIS BOX</u>
Participants = $80 * 1000 = 80,000$ bytes = 80kb
Categories = $50 * 20 = 1,000$ bytes = 1kb
Winners = $20 * 100 = 2,000$ bytes = 2kb [1/2]
Selection and Projection = $14 * 10 = 140$ bytes [1/2]
Join with Category = $64 * 10 = 640$ bytes [1/2]
Projection = $24 * 10 = 240$ bytes [1/2]
Join with Participants = $104 * 10 = 1,040$ bytes [1/2]
Projection = $39 * 10 = 390$ bytes [1/2]

3) (a) (i) Some organisations need multiple security levels. Identify these levels and explain reasons for such classifications.

(03 marks)

<u>ANSWER IN THIS BOX</u>
Top Secret
This type of data could affect many parties.
e.g. customer, organisation
Secret / Confidential
This type of data could cause an effect to a particular party.
e.g. customer
Unclassified (Public)
Represents common information with no effect on any person.

(ii) A database system uses a system log to keep track of all operations on the database. Explain the key usages of this system log.

(04 marks)

ANSWER IN THIS BOX

Database Recovery

A database could fail due to several reasons.

The system log assists to recover the database without losing any completed transaction.

Database Audits

The system log assists in monitoring database user behaviours.

Any unauthorised activities could be detected through this process.

(iii) What are the typical contents of a system log?

(02 marks)

ANSWER IN THIS BOX

Terminal id, user account name,

transaction id, time,

type of operation, object,

old values, new value.

(b) The following is part of an Automated Teller Machine (ATM) database schema belonging to a network of several banks.

```
Bank_Branches(branch-id, bank_name, branch_name, branch_address,
              other_branch_details)
ATM(atm-id, branch-id, location, other_atm_details)
Customers_Cards(card-no, account_id, card_type, date_valid_till,
               daily_withdrawal_limit)
Withdrawals(withdrawal-id, atm-id, card-no, withdrawal_date,
            withdrawal_time, withdrawal_amount)
```

Bank network provides the following access privileges to user groups using their database administrator. Administrator creates user roles and assigns privileges. Branch managers are allowed to pass some of their privileges to their support staff classified as Cashiers.

Customer: View own Withdrawals.

Manager: View transactions belonging to his branch. Add new AMT details.

Cashier: View transactions belonging to the branch.

Administrator: Create users and roles and assigns privileges.

(i) Propose multiple security levels for the above database schema. Identify the data and relations which belong to each security level.

(03 marks)

<u>ANSWER IN THIS BOX</u>	
Top Secret	Customers_Card - card-no, daily_withdrawal_limit
Secret / Confidential	Withdrawals – all data
Public	Bank_Branches – all data
	ATM – all data

(ii) Write SQL statements to create views which would allow a customer to view his withdrawals and branch manager to view branch transactions.

(04 marks)

ANSWER IN THIS BOX

CREATE VIEW Customer-Withdrawals AS

(SELECT * FROM Withdrawals WHERE card-no=\$CUST-CARD-NO)

CREATE VIEW Branch-Withdrawals AS

(SELECT w.* FROM Withdrawals w, ATM a

WHERE a.ATM-id=w.AMT-id AND branch-id=\$BRANCH)

(iii) Write SQL statements to create roles for customers, managers and cashiers, and assign privileges to each of them.

(04 marks)

ANSWER IN THIS BOX

CREATE ROLE Customer

GRANT SELECT ON Customer-Withdrawals TO Customer

CREATE ROLE Manager

GRANT SELECT ON Branch-Withdrawals TO Customer WITH GRANT OPTION

GRANT INSERT ON ATM TO Branch-Manager

CREATE ROLE Cashier

GRANT SELECT ON Branch-Withdrawals TO Cashier

- (iv) A new branch manager B1 had been appointed recently. The database administrator assigns privileges to B1. B1 passes privileges to a cashier CH1 who in turns assigns privileges to a new customer. Write SQL statements to carry out the request as described. Indicate the users who perform these tasks.

(02 marks)

ANSWER IN THIS BOX

Administrator: GRANT Manager TO B1

Manager: GRANT Cashier to CH1

Cashier: GRANT Customer TO C1

- (v) To determine the daily cash allocations for each ATM, the bank wants to provide access to selected statistical data to their statistician. Propose SQL statements to retrieve such data.

(03 marks)

ANSWER IN THIS BOX

CREATE VIEW Stats AS

SELECT atm-id, SUM(withdrawal_amount) AS Day-Withdrawal FROM Withdrawals

GROUP BY atm-id, withdrawal_date

SELECT MAX(Day-Withdrawal), Avg(Day-Withdrawal) FROM Stats

4) (a) In a distributed database, three types of transparencies are possible. Explain each of these types.

(06 marks)

ANSWER IN THIS BOX

Distribution Transparency

It provides freedom for the user from operational details of the network.

Include location and naming transparencies.

Replication Transparency

It allows one to store copies of data at multiple sites for better performance, availability and reliability. User is unaware of these copies.

Fragmentation Transparency

It allows one to introduce horizontal and vertical fragments of data through splitting the original relation. User is unaware of the fragments.

(b) When modelling for Data Warehouses two common multidimensional schemas are the star schema and the snowflake schema. Briefly state what each of them consists of.

(04 marks)

ANSWER IN THIS BOX

Star schema consists of a fact table with a single table for each dimension.

Snowflake schema is a variation on the star schema in which the dimensional tables from a star schema are organised into a hierarchy by normalising them.

(c) Consider the following examination results of a student.

Index	Name	Year	Subject	Grade
0101	AS Perera	2005	IT101	B
			IT102	D
		2006	IT102	C

Write XML syntax to represent this data as an XML document.

(06 marks)

ANSWER IN THIS BOX

```

<student-results>
  <index>0101</index>
  <name>AS Perera</name>
  <year-results>
    <year>2005</year>
    <result>
      <subject>IT101</subject>
      <grade>B</grade>
    </result>
    <result>
      <subject>IT102</subject>
      <grade>D</grade>
    </result>
  </year-results>
  <year-results>
    <year>2006</year>
    <result>
      <subject>IT102</subject>
      <grade>C</grade>
    </result>
  </year-results>
</student-results>

```

(d) SQL standard includes several components such as SQL/CLI and SQL/PSM. Describe what these two components consist of.

(04 marks)

ANSWER IN THIS BOX

SQL/CLI – Call Level Interface provides rules which allow execution of application code without providing source code and avoid pre-processing them. It contains routines for tasks like connecting to SQL server.

SQL/PSM – Persistent Stored Modules specify facilities for participating an application between a client and a server. The goal is to enhance performance by minimizing network traffic.

(e) What are the five types of knowledge produced from data mining?

(05 marks)

ANSWER IN THIS BOX

Association rules – These rules correlate the presence of a set of items with another range of values for another set of variables.

Classification hierarchies – The goal is to work from an existing set of events or transaction to create a hierarchy of classes.

Sequential patterns – A sequence of actions or events is sought. Detection of sequential patterns is equivalent to detecting associations among events with certain temporal relationships.

Pattern within time series – Similarities can be detected within positions of a time series of data, which is a sequence of data taken at regular intervals.

Clustering – A population of events or items can be partitioned into sets of similar elements.

Index No: