University of Ottawa
CSI 3317A, 3317B, and 3717 – Final Exam
Professor(s): Iluju Kiringa, and Michael Bennett

December 19, 2003
14:h00-17h00
Duration: 3 hrs

Closed book; no aid allowed, except one double-sided A4 “cheat sheet”

Family name: _____

First name: _____

Student number: _____

There are 17 questions and a total of 142 points.

This exam must contain 30 pages,
including this cover page.

1 – Multiple Choice I [ALL SECTIONS]	/ 20
2 – Multiple Choice II [3317A+3717 ONLY]	/ 12
3 – Multiple Choice II [3317B ONLY]	/ 12
4 – Concept Definitions [ALL SECTIONS]	/ 8
5 – ER Diagram [3317A+3717 ONLY]	/ 10
6 – Relational Algebra [3317A+3717 ONLY]	/ 6
7 – SQL [3317A+3717 ONLY]	/ 4
8 – Index Matching [3317A+3717 ONLY]	/ 5
9 – Hash-Join [3317A+3717 ONLY]	/ 5
10 – ER Diagram [3317B ONLY]	/ 10
11 – SQL [3317B ONLY]	/ 10
12 – Disks [3317B ONLY]	/ 5
13 – Distributed DBMSs [3317B ONLY]	/ 5
14 – Serializability [ALL SECTIONS]	/ 10
15 – Normal Forms [ALL SECTIONS]	/ 10
16 – ACID Properties [ALL SECTIONS]	/ 5
17 – Two Phase Locking Protocol [ALL SECTIONS]	/ 5
Total	/ 142

READ THESE INSTRUCTIONS FIRST !

1. **Parts of the exam that are marked with "3317A+3717 ONLY"are for CSI3317A+CSI3717 students only;**
2. **Parts of the exam that are marked with "3317B ONLY"are for CSI3317B students only;**
2. **Parts of the exam that are marked with "ALL SECTIONS"are for all CSI3317A, CSI3317B, and CSI3717 students.**

Here are the questions that each section must answer:

- CSI3317A and CSI3717: 1,2,4,5,6,7,8,9,14,15, and one of 16 or 17.
- CSI3317B: 1,3,4,10,11,12,13,14,15, and one of 16 or 17.

The exam consists of 2 parts, 40% Multiple Choice; 60% essay.

The mark allotment is provided for each question.

All work MUST be contained in this Exam book. Use the back of the sheets if necessary.

Note: For multiple choice questions, always choose the most inclusive answer. For example, if you were asked, "who is the prime minister of Canada" and the possible answers were

A. Paul Martin

B. The leader of the liberal party

C. The leader of the party with a majority in the House of Commons

D. The leader of the party asked by the Governor-General to lead the current parliament

You would answer (D) because it is the most inclusive, most general answer.

"University policy states that cheating on an examination is a scholastic offence. The commission of a scholastic offence is attended by academic penalties, which might include expulsion from the program. If you are caught cheating, there will be no second warning".

... Cont'd

1 Multiple Choice I [ALL SECTIONS] — 20 points

Part A — 1 point One of these is **NOT** a characteristic of file-based data systems:

A. Slow access

B. Duplication of data

C. Isolation of data

D. Static queries

Part B — 1 point In the ANSI-SPARC reference model (i.e. the standard reference architecture), if the location of physical data is changed, then

A. The database has to be recompiled

B. The upper layers need to know about the change

C. Nothing happens above the physical layer

D. Only the immediate upper layer needs to know about the change

Part C — 1 point One of these is NOT an aggregate function:

A. COUNT

B. MEDIAN

C. SUM

D. AVG

Part D — 1 point Conceptually, what is the temporal order in which SQL evaluates GROUP BYs?

A. WHERE, HAVING, GROUP BY

B. WHERE, GROUP BY, HAVING

C. HAVING, WHERE, GROUP BY

D. Is irrelevant

... Cont'd

Part E — 1 point Which of the following is **NOT** an advantage of SQL?

A. portability

B. universal standard

C. many different ways to express the same query

D. both interactive and embedded access

Part F — 1 point What does the HAVING clause do?

A. filters groups subject to some condition

B. filters groups of rows

C. filters groups

D. filters groups of rows with the same column value

Part G — 3 points Suppose we have an SQL relation declared by

```
CREATE TABLE FOO(
    name VARCHAR(50) PRIMARY KEY,
    salary INT
    CHECK(salary <= (SELECT AVG(salary) FROM FOO)) );
```

Initially, the Contents of FOO is:

name	salary
'Sally'	10000
'Joe'	20000
'Sue'	30000

We now try to execute the following sequence of modifications, one at a time:

```
INSERT INTO FOO VALUES('Fred', 12000);
UPDATE FOO SET salary = 20000 WHERE name = 'Sue';
INSERT INTO FOO VALUES('Sally', 13000);
DELETE FROM FOO WHERE name = 'Joe';
```

At the end of these statements, the sum of the salaries over all the tuples then in FOO is:

A. 42,000

B. 62,000

C. 65,000

D. 72,000

... Cont'd

Part H — 3 points Suppose we wish to constrain the data so that in no department can the employees have a total salary greater than $1,000,000. The following is a framework for an assertion that will enforce this constraint.

```
CREATE ASSERTION cheap CHECK( NOT EXISTS(Q));
```

Which query Q best enforces this constraint?

A.
```
SELECT * FROM Emps WHERE SUM(salary) > 1000000
```

B.
```
SELECT dept, SUM(salary) FROM Emps GROUP BY dept
```

C.

```
SELECT SUM(salary)
FROM Emps, Managers
WHERE id = mgr
GROUP BY Emps.dept
HAVING SUM(salary) > 1000000
```

D.

```
SELECT dept
FROM Emps
GROUP BY dept
HAVING SUM(salary) > 1000000
```

Part I — 3 points This question is based on a relation `Enroll(SID,CID,term,grade)` which stores the academic records of students in a university database. For example, a tuple `(123, 'SE470b', 'Winter 2003', 9.0)` represents the fact that student `123` took `SE470b` in Winter 2003 and received a grade of `9.0` (which is an A).

Someone in Registrar's Office complained that he obtained different results when repeating the same query inside one transaction:

```
T1:
SELECT AVG(grade) FROM Enroll;     Q1
SELECT AVG(grade) FROM Enroll;     Q2
COMMIT;
```

Which one of the following actions is sufficient and necessary to ensure that Q1 and Q2 always return the same result?

... Cont'd

A. Set transaction level for all transactions, including T1, at SERIALIZABLE

B. Set transaction level for all transactions, including T1, run at REPEATABLE READ

C. Set transaction level for T1 at SERIALIZABLE

D. Set transaction level for T1 at REPEATABLE READ

Part J — 1 point Which of these is **NOT** an aspect of database transactions?

A. RAID

B. wait-for graphs

C. Two-phase commit

D. locks

Part K — 1 point Which of these problems can occur in a multi-user environment?

A. Lost update

B. Uncommitted data

C. Inconsistent data

D. All of the above

Part L — 1 point 2PL does **NOT** do the following:

A. permits deadlocks

B. permits only serializable schedules

C. releases locks only at commit time

D. releases locks progressively

Part M — 1 point How does BCNF differ from 3rd Normal Form?

A. they are the same

B. BCNF has the property that, for every functional dependency $X \rightarrow Y$, X contains a key; i.e., every determinant is a candidate key

C. 3NF contains no transitive dependency on the primary key

... Cont'd

D. BCNF contains partial dependencies on the primary key

Part N — 1 point The difference between pessimistic and optimistic concurrency control protocols is

A. Pessimistic protocols will not allow conflicts to occur

B. Pessimistic protocols will not allow deadlocks to occur

C. Optimistic protocols assume conflicts are rare

D. All of the above

... Cont'd

2 Multiple Choice II [3317A+3717 ONLY] — 12 points

Part A — 2 points

Which one of the following assertions is **NOT** an advantage of using a DBMS?

A. Application programs are not exposed to the details of data representation and storage.

B. The DBMS schedules concurrent access to the data.

C. Administration of data is not centralized.

D. The DBMS protects the data from the effects of system failures.

E. The DBMS enforces the integrity constraints on stored data.

Part B — 2 points Which one of the following assertions is **NOT true**?

A. A data model is a collection of high-level constructs for describing data.

B. The central construct for describing data in the relational model is the notion of schema.

C. The physical schema specifies the storage details of the data model.

D. The conceptual schema defines the data in terms of the data model.

E. An external schema contains a collection of one or more views on relations in the database.

Part C — 2 points Which one of the following assertions is **true**?

A. The goal of conceptual database design is to produce a logical schema of the database.

B. The goal of logical schema refinement is to produce a physical schema of the database.

C. The file of records is an important abstraction in a DBMS, and is implemented by the file and access methods layer of the DBMS code.

D. A heap file is a file of records that are stored on a heap.

E. Range queries are best suited for hash-based indexes.

Part D — 2 points Which one of the following assertions is **NOT true**?

A. In a packed memory page for fixed-length records, the free space is always at the end of the page.

... Cont'd

B. In an unpacked memory page for fixed-length records, the free space is spread all over the page.

C. In a RAID, redundancy and data striping are combined to increase reliability and performance, respectively.

D. RAID levels that use mirroring maintain identical copies of the data on two different disks.

E. A bitmap is an array of bits for keeping track of free slot information in a packed memory page for fixed-length records.

Part E — 2 points Which one of the following assertions is **true**?

A. In an ISAM tree, only internal pages are affected in updates.

B. In an ISAM tree, we delete any empty leaf page.

C. In an ISAM tree, we do not maintain overflow pages.

D. During deletion in an ISAM tree, we delete all overflow pages.

E. Both ISAM and B+ trees support range queries.

Part F — 2 points Review Question Which one of the following assertions is **true**?

A. In a static hash table, primary pages may be deleted.

B. In an extendible hash table, we may have overflow pages.

C. Static hashing is similar to B+ trees.

D. During deletion in a static hash table, we delete nonempty overflow pages.

E. None of the assertions above is true.

... Cont'd

3 Multiple Choice II [3317B ONLY] — 12 points

Part A — 1 point A good reason to change the physical layer would be this:

A. Groom the disks

B. Add more hardware

C. Defrag the disk

D. All of the above

Part B — 1 point The main purpose of IRDS is to

A. Define a data dictionary

B. Define a standard for integrating data dictionaries

C. Isolate data from schemas

D. Share catalogue information

Part C — 1 point QBE is

A. SQL for Dummies

B. a layer on top of SQL

C. easier to use than SQL directly

D. all of the above

Part D — 1 point A NULL is represented by a

A. zero

B. blank

C. nothing

D. a special reserved symbol

Part E — 1 point A serious performance problem with distributed joins is

... Cont'd

A. The slowness of WAN links

B. Head movement time

C. Replication consistency

D. Non-standardized SQL

Part F — 1 point What is true of a Multidatabase system?

A. It is an IRDS standard

B. Each site retains its own autonomy

C. Each site replicates some portion of the whole

D. Each site is WAN-linked to each other

Part G — 1 point A SAN is a

A. Storage Area Network

B. Secure Application Network

C. Synchronized And Normalized DBMS

D. Strategic Area Network

Part H — 1 point Which is NOT an advantage of a DDBMS

A. Local autonomy

B. Improved reliability

C. Improved performance

D. Security

Part I — 1 point Which of these are (is) (a) disadvantage(s) of fragmentation? **(SELECT ALL)**

A. Usage

B. Efficiency

C. Performance

... Cont'd

D. Integrity

Part J — 1 point A horizontal fragment consists of a subset of

A. The tuples of a relation

B. The attributes of a relation

C. A and B

D. The attributes of a relation with high affinity

Part K — 1 point Which of the following are examples of transparency in DDBMSs? **(SELECT ALL)**

A. Distribution

B. Transaction

C. Performance

D. All of the above

Part L — 1 point The textbook gave an example of timing for different distributed query processing strategies that ranged from 1 second to 2.3 days. The main reason for this was

A. Data communication speed variances

B. The amount of data needed to do the query

C. Bad SQL programming

D. DBMS performance issues

... Cont'd

4 Concept Definitions [ALL SECTIONS] — 8 points

For each one of the following concepts, give **a two-lines**, clear definition in the space provided below these concepts:

1. Serial schedule

2. Transaction

3. Clustered index

4. Functional dependency

5. RAID (Give the idea.)

6. Deadlock

7. Atomicity

8. Conflict serializability

... Cont'd

5 ER Diagram [3317A+3717 ONLY] — 10 points

SITE owns a database that contains information about professors. Of course that information is stored in a database with the following relevant specification:

- Every professor has a SIN; he also has a name, and the year in which he was hired.
- There are tenured and nontenured professors. Tenured professors have an indication of the year in which they got their tenure. Nontenured professors have an indication of the year in which they can apply for tenure.
- Among tenured professors, there are associate professors and full professors. Full professors have an indication of the year in which they were promoted to full professorship.
- Nontenured professors must be expert in a given area.
- All the areas of expertise have an identification, and a description.
- A full professor can advise a nontenured professor on his expertise in a given area. On the other hand, each expertise must have a full professor advising on it.

Draw an ER diagram for this setting. Make sure to indicate all the appropriate constraints.

... Cont'd

6 Relational Algebra [3317A+3717 ONLY] — 6 points

Consider the following relational schema:

$$Suppliers(\underline{sid : \texttt{int}}, sname : \texttt{string}, adress : \texttt{string}, rating : \texttt{real})$$
$$Parts(\underline{pid : \texttt{int}}, pname : \texttt{string}, color : \texttt{string}, fid : \texttt{int})$$
$$Catalog(\underline{sid : \texttt{int}, pid : \texttt{int}}, cost : \texttt{real})$$

The Catalog relation lists the prices for parts by suppliers. Write the following queries in **relational algebra**:

- Find the sids of suppliers who supply every red part or supply every green part.
- Find the pids of parts supplied by every supplier at less than $200.
- Find the names of suppliers who supply a green part that costs less than $50.

... Cont'd

7 SQL [3317A+3717 ONLY] — 4 points

Consider the schema of the previous question.

1. Write the following query in SQL:

"Find the names of suppliers who supply a green part that costs less than $50."

2. Write an SQL statement to create the Catalog relation, including the following integrity constraint expressed using CHECK:

"Every part has a minimum of 2 suppliers and a maximum of of 10 suppliers that supply it ."

... Cont'd

8 Index Matching [3317A+3717 ONLY] — 5 points

Consider the following relational schema of the Suppliers relation:

$$Suppliers(\underline{sid} : \texttt{int}, sname : \texttt{string}, adress : \texttt{string}, rating : \texttt{real})$$

For each of the following indexes, say whether the index matches the given selection conditions. List the primary conjuncts for every match. That is, say **yes** or **no**, and if yes, list the primary conjuncts.

(1) A B+ tree index on the search key $\langle Suppliers.sid \rangle$

(a) $\sigma_{Suppliers.sid>20{,}000}(Suppliers)$

(b) $\sigma_{Suppliers.sid<20{,}000}(Suppliers)$

(c) $\sigma_{Suppliers.sid=20{,}000}(Suppliers)$

(2) A B+ tree index on the search key $\langle Suppliers.sid, Suppliers.rating \rangle$

(a) $\sigma_{Suppliers.sid>20{,}000 \wedge Suppliers.rating<4}(Suppliers)$

(b) $\sigma_{Suppliers.rating<4}(Suppliers)$

(c) $\sigma_{Suppliers.sid=20{,}000}(Suppliers)$

(2) A hash index on the search key $\langle Suppliers.sid, Suppliers.rating \rangle$

(a) $\sigma_{Suppliers.rating<4}(Suppliers)$

(b) $\sigma_{Suppliers.sid>20{,}000 \wedge Suppliers.rating<4}(Suppliers)$

(c) $\sigma_{Suppliers.sid=20{,}000}(Suppliers)$

... Cont'd

9 Hash-Join [3317A+3717 ONLY] — 5 points

Explain how the hash-join algorithm works.

... Cont'd

10 ER Diagram [3317B ONLY] — 10 points

Represent each of the following requirements with an ER diagram:

(a) A company called Perfect Pets runs a number of clinics. A clinic has many staff and a member of staff manages at most one clinic (not all staff manage clinics). Each clinic has a unique clinic number (clinicNo) and each member of staff has a unique staff number (staffNo).

(b) When a pet owner contacts a clinic, the owner's pet is registered with the clinic. An owner can own one or more pets, but a pet can only register with one clinic. Each owner has a unique owner number (ownerNo) and each pet has a unique pet number (petNo).

(c) When the pet comes along to the clinic, it undergoes an examination by a member of the consulting staff. The examination may result in the pet being prescribed with one or more treatments. Each examination has a unique examination number (examNo) and each type of treatment has a unique treatment number (treatNo).

... Cont'd

(d) Represent the complete set of requirements in one ER diagram.

... Cont'd

11 SQL [3317B ONLY] — 10 points

The following tables form part of a database held in a relational DBMS:

$$Hotel(hotelNo, hotelName, city)$$
$$Room(roomNo, hotelNo, type, price)$$
$$Booking(hotelNo, guestNo, dateFrom, dateTo, roomNo)$$
$$Guest(guestNo, guestName, guestAddress)$$

Here, Hotel contains hotel details and hotelNo is the primary key; Room contains room details for each hotel and (roomNo, hotelNo) forms the primary key; Booking contains details of the bookings and (hotelNo, guestNo, dateFrom) forms the primary key; and Guest contains guest details and guestNo is the primary key.

Part A — 2 points What does the following code snippet do? Be very precise.

```
CREATE DOMAIN RoomType AS CHAR(1)
       CHECK(VALUE IN ('S', 'F', 'D'));
CREATE DOMAIN HotelNumbers AS HotelNumber
       CHECK(VALUE IN (SELECT hotelNo FROM Hotel));
CREATE DOMAIN RoomPrice AS DECIMAL(5, 2)
       CHECK(VALUE BETWEEN 10 AND 100);
CREATE DOMAIN RoomNumber AS VARCHAR(4)
        CHECK(VALUE BETWEEN '1' AND '100');

CREATE TABLE Room(
        roomNo RoomNumber NOT NULL,
        hotelNo HotelNumbers NOT NULL,
        type  RoomType NOT NULL DEFAULT 'S'
        price RoomPrice NOT NULL,
        PRIMARY KEY (roomNo, hotelNo),
        FOREIGN KEY (hotelNo) REFERENCES Hotel
              ON DELETE CASCADE ON UPDATE CASCADE);
```

... Cont'd

Part B — 2 points Create a view containing the hotel name and the names of the guests staying at the hotel.

... Cont'd

Part C — 2 points What does the following code snippet do?

```
GRANT ALL PRIVILEGES ON HotelData
TO Manager, Director WITH GRANT OPTION;
GRANT ALL PRIVILEGES ON BookingOutToday
TO Manager, Director WITH GRANT OPTION;
```

... Cont'd

Part D — 4 points Which of the following is semantically correct?

(a)
```
SELECT r.type, r.price
FROM Room r, Hotel h
WHERE r.hotel_number = h.hotel_number AND
h.hotel_name = 'Grosvenor Hotel' AND
                                    r.type > 100;
```

(b)
```
SELECT r.roomNo, h.hotelNo
FROM Hotel h, Booking b, Room r
WHERE h.hotelNo = b.hotelNo AND h.hotelNo = 'H21' AND
   b.roomNo = r.roomNo AND type = 'S' AND b.hotelNo = 'H22';
```

(c)
```
SELECT r.roomNo, r.type, r.price
FROM Room r, Booking b, Hotel h
WHERE r.roomNo = b.roomNo AND b.hotelNo = h.hotelNo AND
h.hotelName = 'Grosvenor Hotel' AND r.price > 100;
```

... Cont'd

12 Disks [3317B ONLY] — 5 points

Suppose that you had a 4-tuple address $\langle diskunit, cylinder, head, record \rangle$ for your disk system and its size is $\langle 4, 8, 4, 8 \rangle$. Assuming the normal 8K physical allocation block size, answer the following:

(a) How many blocks can be addressed if the system was fully populated?

(b) What is the maximum size of data that a cylinder can hold?

(c) What is the slow-down ratio of a cylinder read compared with the worse case single block multiple read? Assume 100 ms for the arm move.

13 Distributed DBMSs [3317B ONLY] — 5 points

Suppose that we have a distributed database system. What things must you consider in replicating important data?

... Cont'd

14 Serializability [ALL SECTIONS] — 10 points

Part A — 4 points Consider the following schedules.

$$S_1 = R_2(X), R_1(Y), W_2(X), R_3(X), W_1(Y), W_3(X), R_2(Y), W_2(Y), C_1, C_2, C_3$$
$$S_2 = R_2(X), R_1(Y), W_2(X), R_2(Y), R_3(X), W_1(Y), W_3(X), W_2(Y), C_1, C_2, C_3$$

For each one of them, tell whether it is conflict serializable or not; and if not say why.

Part B — 6 points For each one of the following schedules:

$$S_1 = R_1(X), R_2(X), R_3(Y), W_1(X), R_2(Z), R_2(Y), W_2(Y), W_1(Z), C_1, C_2, C_3$$
$$S_2 = R_1(X), W_1(Y), R_2(Y), W_2(Z), R_3(Z), W_3(X), C_1, C_2, C3$$
$$S_3 = W_3(X), R_1(X), W_1(Y), R_2(B), W_2(Z), R_3(Z), C_1, C_2, C3$$

Answer the questions below:

(1) What is the precedence graph of the schedule?

(2) Give an equivalent serial schedule for any conflict serializable schedule.

... Cont'd

15 Normal Forms [ALL SECTIONS] — 10 points

The table shown below displays the details of the roles played by actors/actresses in films.

fNo	fTitle	dirNo	dName	actorNo	actorName	role	timeOnScreen
F1100	Happy Days	D101	Jim Alan	A1020	Sheila Toner	Jean	15.45
F1100	Happy Days	D101	Jim Alan	A1222	Peter Watt	Tom	25.38
F1100	Happy Days	D101	Jim Alan	A1020	Sheila Toner	Silvia	15.45
F1109	Snake Bite	D076	Sue Ramsay	A1567	Steven McDonald	Tim	19.56
F1109	Snake Bite	D076	Sue Ramsay	A1222	Peter Watt	Archie	10.44

(a) Describe why the table shown above is not in first normal form (1NF).

(b) The table shown above is susceptible to update anomalies. Provide examples of how insertion, deletion, and modification anomalies could occur on this table.

(c) Identify the functional dependencies represented by the table shown above. State any assumptions you make about the data shown in this table (if necessary).

(d) Using the functional dependencies identified in part (c), describe and illustrate the process of normalization by converting the table shown above to third normal form. Identify the primary and foreign keys in your relations.

... Cont'd

16 ACID Properties [ALL SECTIONS] — 5 points

Consider the ACID properties. In one page, describe why they are important for transactions.

... Cont'd

17 Two Phase Locking Protocol [ALL SECTIONS] — 5 points

Explain how the two-phase locking protocol works.

... End of Final Exam