THE UNIVERSITY
OF QUEENSLAND
AUSTRALIA

This exam paper must not be removed from the venue

Venue   _____

Seat Number   _____

Student Number |__|__|__|__|__|__|__|__|

Family Name   _____

First Name   _____

School of Information Technology and Electrical Engineering

EXAMINATION

Semester Two Final Examinations, 2019

INFS7903 Relational Database Systems

This paper is for St Lucia Campus students.

Examination Duration: 120 minutes

Reading Time: 10 minutes

Exam Conditions:

This is a Central Examination

This is a Closed Book Examination - no materials permitted

During reading time - write only on the rough paper provided

This examination paper will be released to the Library

Materials Permitted In The Exam Venue:

(No electronic aids are permitted e.g. laptops, phones)

Calculators - Casio FX82 series or UQ approved (labelled)

Materials To Be Supplied To Students:

None

Instructions To Students:

Please answer all questions on the examination paper.

Total marks: 100 (to be scaled down to 60)

Additional exam materials (eg. answer booklets, rough paper) will be provided upon request.

For Examiner Use Only

Question	Mark

Total _____

Question 1 [6 marks] A *COURSE* relation contains the following fields:

`CID: integer, CName: string, DeptNo: integer, NEnrol: integer`

Assume that an integer occupies 4 bytes and a character occupies 1 byte. Consider the INSERT statement below:

`INSERT INTO COURSE VALUES (7903, 'Database', 1, 152);`

1.1) [2 marks] Assume that **fixed-length record organization** is used, and the field *CName* is stored using 20 characters. How many bytes will the above inserted record occupy? Show your calculation process and result.

1.2) [2 marks] Assume that **variable-length record organization** is used, and an integer is used to store the field length of *CName*. How many bytes will the above inserted record occupy? Show your calculation process and result.

1.3) [2 marks] Explain the advantages and disadvantages of variable-length record organization compared to fixed-length record organization.

Question 2 [10 marks] Consider the *COURSE* relation defined in Question 1 and the fixed-length record organization used in Question 1.1. The *COURSE* relation contains 10,000 records and it is stored on a disk with the following configuration:

- Block size (B) = 512 bytes
- Block pointer (P) = 12 bytes long

2.1) [1 mark] Suppose the *COURSE* file is ordered by the non-key field *DeptNo* and we want to construct a single-level index on this field. Which index is to be used, primary index, clustering index, or secondary index?

2.2) [4 marks] Assume there are 100 distinct values of *DeptNo*, and the *COURSE* records are evenly distributed among these values. How many blocks are required to store the above index? Show your calculation process and result.

2.3) [5 marks] Now we extend the above single-level index to a multi-level index. Assume that block anchors are used for the *COURSE* file (i.e., every new value of *DeptNo* starts at the beginning of a new block) and data records with the same *DeptNo* are stored in adjacent disk blocks. How many block accesses are needed to retrieve all *COURSE* records having a specific *DeptNo* value using the multi-level index? Show your calculation process and result.

Question 3 [8 marks] Consider the following *Electronics* relation. Suppose the field *Item* has two distinct values: 'computer' and 'phone'; the field *Location* has four distinct values: 'Sydney, 'Brisbane', 'Melbourne', and 'Canberra'.

RID	Item	Location	Sales
R1	computer	Sydney	882
R2	computer	Brisbane	968
R3	computer	Melbourne	746
R4	computer	Canberra	825
R5	phone	Sydney	89
R6	phone	Brisbane	38
R7	phone	Melbourne	43
R8	phone	Canberra	14

3.1) [2 marks] If a bitmap index is to be created on field *Location*, what is the total size of that index in bits? Show your calculation process and result.

3.2) [2 marks] Show the bitmap corresponding to the value 'Brisbane'.

3.3) [4 marks] Given the bitmap index on *Item* and *Location* respectively, explain how to find the *Sales* information of 'computer' in 'Brisbane'.

Question 4 [6 marks] Consider the following B+ tree:
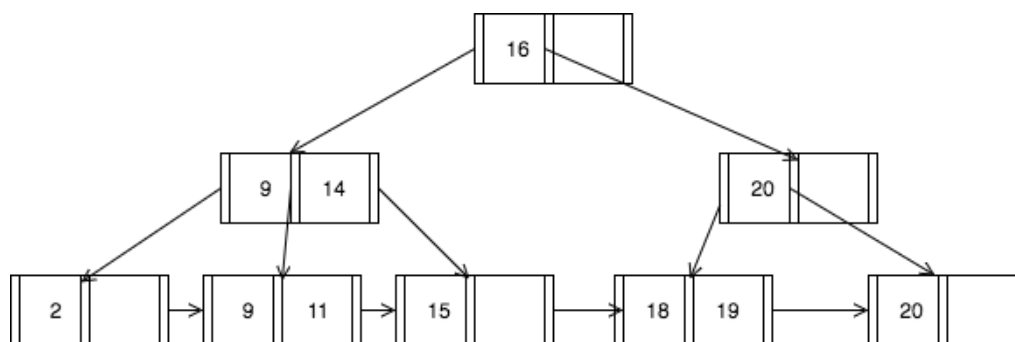
4.1) [3 marks] How many tree nodes need to be accessed to retrieve all records with key larger than 17? Explain your answer.

4.2) [3 marks] What would be the total number of nodes after inserting the key '10'? Explain your answer.

Question 5 [10 marks] Given two relations R1 and R2, where R1 contains N1 tuples and R2 contains N2 tuples (N2 > N1 > 0), what are the minimum and maximum possible sizes (in number of tuples) for the resulting relation produced by each of the following relational algebra expressions? Explain your answer (i.e., in what situation would that minimum or maximum size happen).

5.1) [2 marks] R1 $\cup$ R2 (Union)

5.2) [2 marks] R1 $\cap$ R2 (Intersection)

5.3) [2 marks] R2 – R1 (Difference)

5.4) [2 marks] $\sigma_{A=5}$ (R1), assuming A is a field in R1 (Selection)

5.5) [2 marks] π_A (R1), assuming A is a field in R1 (Projection)

Question 6 [10 marks] Consider two relations R1 and R2 stored on a disk with block size = 1000 bytes. R1 contains 200,000 records with each record occupying 50 bytes. R2 contains 10,000 records with each record occupying 20 bytes. Assume that the size of available memory is 52 blocks. Nested-loop join is used to answer R1 ⋈ R2.

6.1) [2 marks] Which relation should be used as the outer relation (i.e., in the outer loop) in order to speed up the join? Explain your answer.

6.2) [4 marks] Let R1 be the outer relation and R2 be the inner relation. Estimate the number of block accesses using the **page-oriented nested-loop join** strategy. Show your calculation process and result.

6.3) [4 marks] Let R1 be the outer relation and R2 be the inner relation. Estimate the number of block accesses using the **block nested-loop join** strategy. Show your calculation process and result.

Question 7 [12 marks] Consider the following relations:
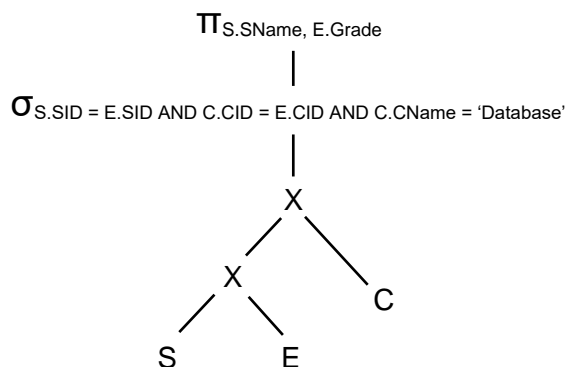
```
Student (SID, SName, Age, GPA)
Course (CID, CName, Lecturer)
Enrollment (SID, CID, Grade)
```

Given the following SQL query:

```
SELECT S.SName, E.Grade
FROM Student S, Course C, Enrollment E
WHERE S.SID = E.SID AND C.CID = E.CID AND C.CName = 'Database';
```

The initial query tree is illustrated as below:

$\pi_{S.SName,\ E.Grade}$

|

$\sigma_{S.SID\ =\ E.SID\ AND\ C.CID\ =\ E.CID\ AND\ C.CName\ =\ \text{'Database'}}$

|

X

/ \

X   C

/ \

S   E

7.1) [3 marks] Based on the above initial query tree, show the equivalent query tree after pushing down selection operators.

7.2) [2 marks] Based on the query tree obtained in Question 7.1, show the equivalent query tree after converting cross-products into joins.

7.3) [2 marks] Based on the query tree obtained in Question 7.2, show the equivalent query tree after rearranging leaf nodes so as to execute the most restrictive selection operators first.

7.4) [5 marks] Based on the query tree obtained in Question 7.3, show the equivalent query tree after pushing down projection operators.

Question 8 [8 marks] A transaction in a relation database system must maintain the ACID properties in order to ensure its accuracy, completeness, and data integrity.

8.1) [4 marks] Name and briefly explain each of the **ACID** properties.

8.2) [4 marks] We have discussed various techniques in this course, such as integrity constraints, storage management, indexing, query optimization, concurrency control, recovery, etc. For each of the ACID properties, explain which technique can be used to guarantee that property.

Question 9 [10 marks] Consider two transactions T1, T2 and two data items X, Y.

`T1 = r(X); w(X); r(Y); w(Y)`

`T2 = r(X); w(X)`

Given the following schedule of interleaved operations from T1 and T2:

`r1(X); r2(X); w1(x); r1(Y); w2(x); w1(Y)`

9.1) [2 marks] What type of **anomaly** occurs in this schedule? Explain your answer.

9.2) [2 marks] Two-Phase Locking (2PL) is widely used for concurrency control in a relational database system. Briefly describe the **basic** 2PL protocol.

9.3) [4 marks] For each of T1 and T2, insert all the **lock** and **unlock** instructions to make the transaction satisfy the basic 2PL protocol.

9.4) [2 marks] Timeout is a mechanism for handling deadlocks. Explain the advantages and disadvantages of **short timeout** compared to **long timeout**.

Question 10 [10 marks] For each of the following schedules: 1) construct a **precedence graph**, 2) determine if the schedule is **conflict serializable**, and 3) show the equivalent serial schedule.

10.1) [5 marks] `r1(X); w1(X); r3(X); r2(X); w3(X)`

10.2) [5 marks] `r3(X); r2(X); w3(X); r1(X); w1(X)`
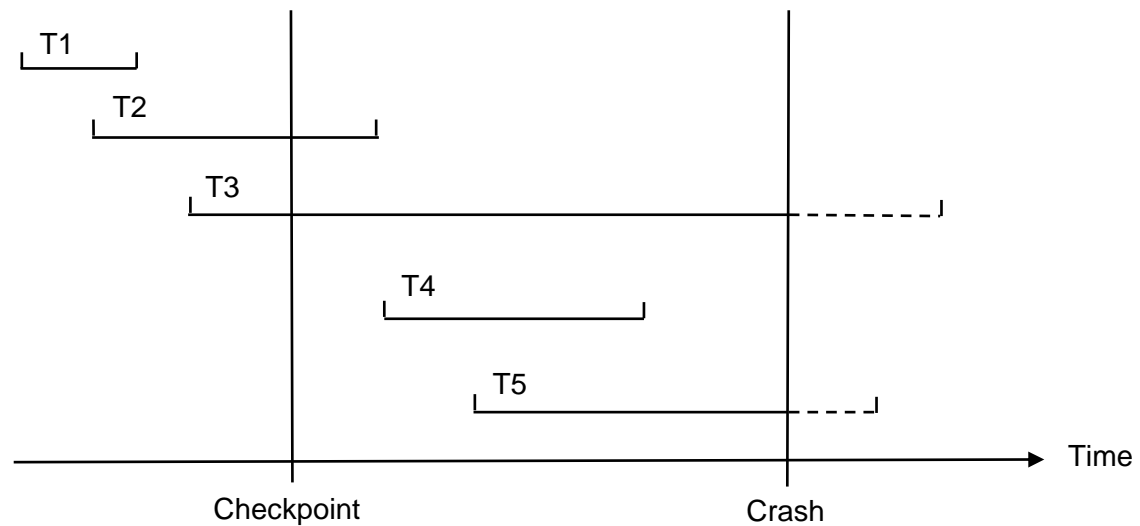
Question 11 [10 marks] Database recovery is necessary when the system encounters some failure, such as transaction failure, system failure, media failure, etc.

11.1) [2 marks] Briefly explain the **write-ahead logging (WAL)** protocol.

11.2) [2 marks] If a **steal**/**no-force** buffer management policy is used, which recovery operations are needed for system recovery? Justify your answer.

11.3) [1 mark] Briefly explain the meaning of a **checkpoint**.

11.4) [5 marks] Consider the following transactions T1, T2, T3, T4, T5, where the two endpoints represent the start time and commit time of a transaction, respectively. For example, T5 starts after the checkpoint and it is not committed before the system crash. For each of the five transactions, which recovery operations are needed? Justify your answer.

END OF EXAMINATION