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, 2020

INFS7903 Relational Database Systems

This paper is for St Lucia Campus students.

Examination Duration: 120 minutes

Reading Time: 10 minutes

Exam Conditions:

Set start and completion time for all students e.g. a 2 hour exam, starts at 8am, ends at 10am

Paper-based exam (on-campus exam only)

This is a Closed Book examination - no written materials permitted

Casio FX82 series or UQ approved (labelled)

Materials Permitted In The Exam Venue:

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

None

Materials To Be Supplied To Students:

None

Instructions To Students:

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

Please answer all questions on the examination paper

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

For Examiner Use Only

Question	Mark

Total ________

Question 1 [5 marks] An *Employee* relation contains the following fields:

`EID: integer, Ename: string, Email: string, Salary: real`

Both *EID* and *Email* are unique fields and can be regarded as candidate keys. In your opinion, which field is a more suitable primary key for the *Employee* relation? List at least four reasons to justify your answer.

Question 2 [7 marks] Consider the following relations in a *Hotel* database:

`Hotel (HotelNo, HotelName, Address)`

`Room (RoomNo, HotelNo, Type, Capacity, Price)`

`Booking (RoomNo, HotelNo, Date, NumberOfGuests)`

2.1) [3 marks] Assume the *Type* field in relation *Room* is defined as *CHAR(6)*, and its value must be one of ‘Single’, ‘Double’, or ‘Family’. Write the SQL statement to CREATE DOMAIN RoomType that enforces this constraint.

2.2) [4 marks] Write the SQL statement to define an assertion that ensures the number of guests in a room cannot exceed the room capacity.

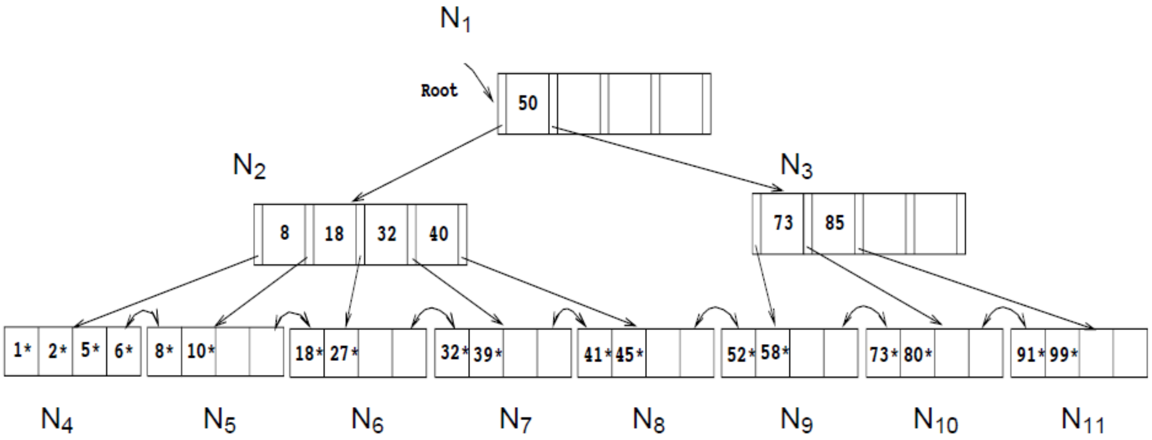
Question 3 [4 marks] There are typically two methods to implement the multiple-disk organization: **data partitioning** and **data mirroring**. Briefly explain each method, and analyse the pros and cons of data mirroring compared with data partitioning.

Question 4 [5 marks] Consider a file which has N = 30,000 *Movie* records. A B+ tree index is constructed on the primary key *MovieID* (integer, 4 bytes), and stored on a disk with the following configuration:

- Block size (B) = 500 bytes
- Block pointer size (P) = 6 bytes

Each tree node is approximately 60% full on average. What is the height of the B+ tree? Show your calculation process and result.

Question 5 [9 marks] Consider the B+ tree as shown in the figure below, where the tree nodes are labelled as N_1, N_2, …, N_{11}. Assume the following rule applies for redistributing keys after a leaf node split: **Two keys** stay in the old leaf node and the remaining keys move to a new leaf node.



5.1) [4 marks] What is the minimum number of tree nodes that must be visited to answer the query: "Get all records with the key greater than 30 and less than 51"? List all the visited tree nodes.

5.2) [5 marks] Show the updated B+ tree after inserting an entry with key "3". For simplicity, you can show only the updated or newly-created tree nodes.

Question 6 [8 marks] Consider the following *Student* relation with N = 500 records:

```
Student (SID, SName, Email, Age, Gender, GPA)
```

SID is the primary key. The *Gender* field has two distinct values: Female and Male, and the *GPA* field has seven distinct values: 1, 2, 3, 4, 5, 6, 7.

6.1) [2 marks] What is the selectivity of "*SID* = 1234"? Show your calculation process and result.

6.2) [2 marks] Assume that the *Student* records are evenly distributed on the *Gender* field. What is the selectivity of "*Gender* = Female"? Show your calculation process and result.

6.3) [4 marks] Assume that *Student* records are distributed as follows on the *GPA* field: 10% with *GPA* = 1; 10% with *GPA* = 2; 15% with *GPA* = 3; 20% with *GPA* = 4; 30% with *GPA* = 5; 10% with *GPA* = 6; 5% with *GPA* = 7. What is the estimated number of *Student* records satisfying "*GPA* > 4"? Show your calculation process and result.

Question 7 [10 marks] Consider two relations R(A, B, C) and S(D, E, A). Field A is the primary key of relation R, and field D is the primary key of relation S. Field A in relation S is a foreign key that references relation R. R and S are stored on a disk with block size = 1000 bytes. Relation R contains 200,000 records with each record occupying 50 bytes. Relation S contains 10,000 records with each record occupying 20 bytes. Consider R * S (natural join). Let S be the outer relation and R be the inner relation.

7.1) [4 marks] Assume that the size of available memory is 52 blocks. Estimate the number of block accesses using the **block nested-loop join** strategy. Show your calculation process and result.

7.2) [6 marks] Assume that R is unsorted, and a multi-level index is constructed on the primary key A. Each index entry occupies 20 bytes. Estimate the number of block accesses using the **single-loop join** strategy with the index. Show your calculation process and result.

Question 8 [13 marks] Consider the following relations:

```
Employee (EID, EName, Age, Salary)

Department (DID, DName, Budget, Manager)

Works (EID, DID, DateFrom, DateTo)
```

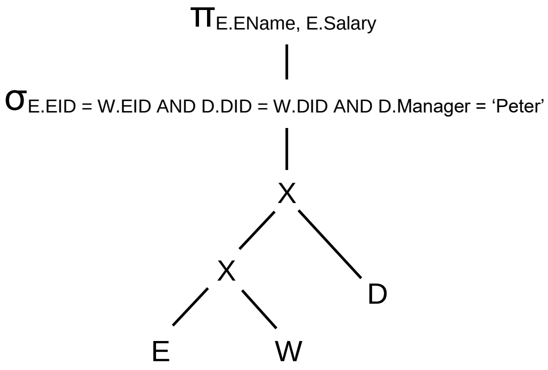
Given the following SQL query:

```
SELECT E.EName, E.Salary

FROM Employee E, Department D, Works W

WHERE E.EID = W.EID AND D.DID = W.DID AND D.Manager = ‘Peter’;
```

The initial query tree is illustrated as below:


$\pi_{E.EName,\ E.Salary}$
$\sigma_{E.EID\ =\ W.EID\ AND\ D.DID\ =\ W.DID\ AND\ D.Manager\ =\ ‘Peter’}$
X
X
D
E
W

8.1) [3 marks] Based on the above initial query tree, show the equivalent query tree after pushing down selection operators.

8.2) [2 marks] Based on the query tree obtained in Question 8.1, show the equivalent query tree after converting cross-products into joins.

8.3) [2 marks] Based on the query tree obtained in Question 8.2, show the equivalent query tree after rearranging leaf nodes so as to execute the most restrictive selection operators first.

8.4) [6 marks] Based on the query tree obtained in Question 8.3, show the equivalent query tree after pushing down projection operators.

Question 9 [8 marks] Query optimization is very important in a DBMS.

9.1) [3 marks] What are the main components of a **query execution plan**?

9.2) [5 marks] Briefly describe the **cost-based query optimization**, and list at least three cost factors typically considered in cost-based query optimization.

Question 10 [15 marks] Consider concurrency control and recovery techniques used in a relational database system.

10.1) [3 marks] Briefly explain each of the following **anomalies** that might occur during transaction execution:

- Lost update
- Dirty read
- Unrepeatable read

10.2) [4 marks] Two-Phase Locking (2PL) is widely used for concurrency control in a DBMS. Briefly explain the **basic 2PL** protocol.

10.3) [2 marks] Timeout is a mechanism for handling deadlocks. Briefly explain the pros and cons of **short timeout** compared with **long timeout**.

10.4) [4 marks] Briefly explain the **write-ahead logging (WAL)** protocol.

10.5) [2 marks] What are the problems of a **no-steal/force** buffer management policy in terms of system efficiency?

Question 11 [6 marks] Consider the following schedule that is generated by some concurrency control protocol for executing two transactions T1 and T2:

```
S = T1:W(X), T2:R(Y), T1:R(Y), T2:R(X), T1:Commit, T2:Commit
```

For each of the following concurrency control protocols:

- State if the protocol allows schedule S, that is, allows the actions to occur in exactly the order shown in schedule S;
- Clearly explain the reason why schedule S is allowed or not allowed under that protocol.

11.1) [2 marks] Under the **Basic 2PL** protocol

11.2) [2 marks] Under the **Strict 2PL** protocol

11.3) [2 marks] Under the **Conservative 2PL** protocol

Question 12 [10 marks] For each of the following schedules:

- Construct a **precedence graph**;
- Determine if the schedule is **conflict serializable**;
- Show the equivalent serial schedule.

12.1) [5 marks] `R1(X); W1(X); R3(X); R2(X); W3(X)`

12.2) [5 marks] `R3(X); R2(X); W3(X); R1(X); W1(X)`

END OF EXAMINATION