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 One Final Examinations, 2018

INFS7901 Database Principles

This paper is for St Lucia Campus students.

Examination Duration: 90 minutes

Reading Time: 10 minutes

Exam Conditions:

This is a School Examination

This is a Closed Book Examination - specified 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 - Any calculator permitted - unrestricted

Two sheets of typed notes double sided are permitted

Materials To Be Supplied To Students:

None

Instructions To Students:

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

For Examiner Use Only

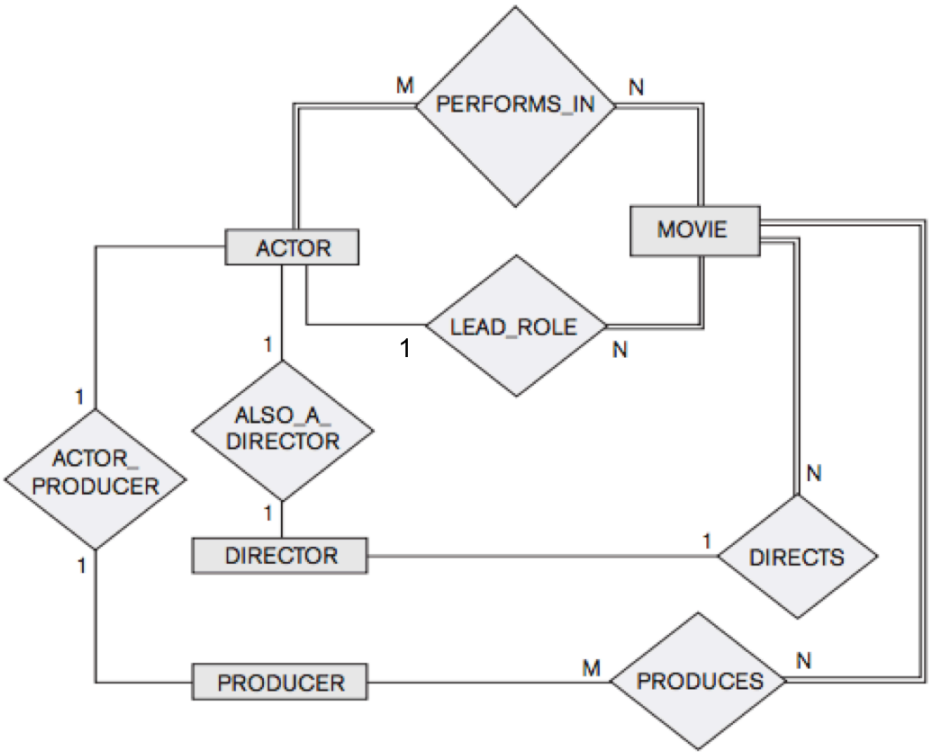
Question	Mark

Total _____

Module 1

Question 1. (3 marks) Entity-Relationship diagrams

Consider the ER schema for the MOVIES database give below. Assume that MOVIES is a populated database. Respond to the following with circling True, False or Maybe. Assign a response of Maybe to statements that while explicitly not shown to be True, cannot be proven False based on the schema as shown.



a) (1 mark) All of the actors in this database have been in at least one movie.
True False Maybe

b) (1 mark) There are some actors who have acted in more than ten movies.
True False Maybe

c) (1 mark) A movie can have a maximum of one lead actor
True False Maybe

Question 2. (6 Marks) **Functional Dependencies and Normal Forms**

Consider the following relational schema and associated function dependencies:

R [CallNo, BookTitle, Price, Publisher, Address, Author, DatePublished, MemberID, Name, Class, Entitlement, DateOfIssue]

- CallNo -> BookTitle, Price, Publisher, Author, DatePublished
- Publisher -> Address
- MemberID -> Name, Class, Entitlement, CallNo
- MemberID -> DateOfIssue
- Class -> Entitlement

a) (1 mark) Find a candidate key for R

b) (1 mark) Determine the closure of MemberID

c) (4 marks) Decompose R into BCNF relations

Question 3. (6 Marks) **SQL**

The following relations keep track of airline flight information. Note that the Employees relation describes pilots and other kinds of employees as well; every pilot is certified for some aircraft, and only pilots are certified to fly.

- Flights (flno, from, to, distance, departs, arrives, price)
- Aircraft (aid, aname, cruisingrange)
- Certified (eid, aid)
- Employees (eid, ename, salary)

Write each of the following queries in SQL

a) (3 marks) For each pilot who is certified for more than three aircrafts, find their eid and the aircraft with the largest cruising range for which they are certified.

b) (3 marks) Find the names of pilots whose salary is less than the price of the cheapest route from Los Angeles to Honolulu.

Module 2

Question 4. (6 Marks) **Asymptotic Analysis**

Answer the following questions

a) (2 marks) Assume that the following two functions show the running time of two algorithms solving the same problem. Which one would you recommend to be used for large values of n? Briefly justify your answer.

Algorithm A's running time = $300n + 100n^{0.1}$

Algorithm B's running time = $n^2 + 10 \log n$

b) (4 marks) Determine the running time of the following code in terms of big-O notation. Show all of your work

```
def secretFunction(aList):
    for i in range(1, len(aList)):
        tmp = aList[i]
        k = i
        while k > 0 and tmp < aList[k - 1]:
            aList[k] = aList[k - 1]
            k -= 1
        aList[k] = tmp
```

Question 5. (6 Marks) **Sorting**

a) Determine whether the following statements are True or False. Briefly justify your answer

I. (1 mark) On average, MergeSort asymptotically runs faster than quicksort

II. (2 marks) The best-case scenario for both insertion sort and quicksort occur when the initial array of input is already sorted.

b) (2 marks) As part of a group project, a student codes an implementation of a certain sorting algorithm. The student tests the algorithm many times with randomly generated arrays of many different sizes to ensure that the algorithm is implemented correctly. However, when the algorithm is put to use it runs much slower that in did in testing, even for arrays of comparable sizes.

Based on this information, which sorting algorithm do you think is being used and why do you think it is running much slower? Briefly justify your answer

c) (1 mark) Suppose you have the following list of numbers to sort:

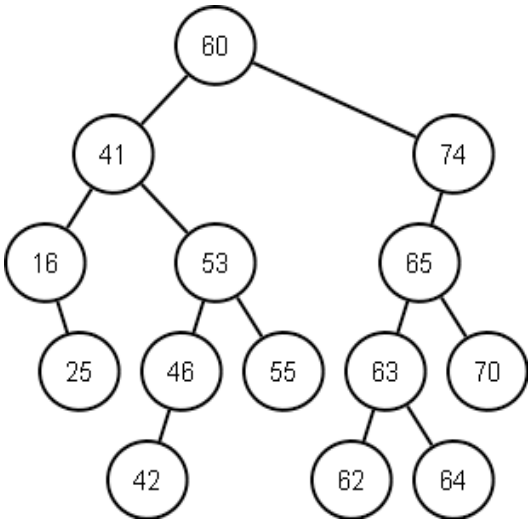
[15, 5, 4, 18, 12, 19, 14, 10, 8, 20]

which list represents the partially sorted list after three complete passes of insertion sort?

- ☐ [4, 5, 12, 15, 14, 10, 8, 18, 19, 20]
- ☐ [15, 5, 4, 10, 12, 8, 14, 18, 19, 20]
- ☐ [4, 5, 15, 18, 12, 19, 14, 10, 8, 20]
- ☐ [15, 5, 4, 18, 12, 19, 14, 8, 10, 20]

Question 6. (8 Marks) **Binary Search Trees**

This question makes reference to the following Binary Search Tree:



a) (1 mark) What is the root of this tree?

b) (1 mark) Which node is the successor of the node with value 41?

c) (1 mark) Which nodes are leaves?

d) (3 marks) Draw the tree resulting from deletion of node with value 60, assuming that we are using the node's successor to replace it.

e) (2 marks) In a balanced Binary tree, what would be the running time of finding the minimum value in the tree? Briefly describe your answer

Note: A tree is balanced if the heights of its subtrees differ by no more than 1. (That is, if the subtrees have heights h1 and h2, then $|h1 - h2| \leq 1$.

Module 3

Question 7. (10 Marks) **Hashing**

a) (4 marks) Given a hashtable table of size 7, using open addressing with the hash function h(x) = h%size, at which indexes would there be a collision using linear probing when the following keys are being inserted? What would the final hash table be?

3, 8, 13, 20, 9, 22

b) (6 marks) Respond to the following with circling True or False

I. Chaining is a method in which there are a list of entries for a single index.
True False

II. Double hashing is vulnerable to secondary clustering.
True False

III. (Assuming no collision) All properly implemented hashing has O(1) add and find functions.
True False

IV. A hash table must be sorted to function correctly.
True False

V. Open-addressing only works up to a load factor of 1.
True False

VI. If load factor < 1, quadratic probing can always find an empty slot
True False

Question 8. (6 Marks) **Indexing**

For this question, we will be using the schema that was introduced in Question 3:

- Flights (flno, from, to, distance, departs, arrives, price)
- Aircraft (aid, aname, cruisingrange)
- Certified (eid, aid)
- Employees (eid, ename, salary)

Suppose each of the following queries occur frequently. For each one, decide which attribute, if any, should be indexed and whether each index should be a clustered index or an unclustered index. Assume that both B+ trees and hashed indexes are supported by the DBMS and that both single- and multiple-attribute index

a) (2 marks) Find name and cruising range of aircrafts based on their aid.

b) (2 marks) Find the eids of employees who make more than $50,000.

c) (2 marks) Find flights that board from New York.

Question 9. (9 Marks) **Relational Algebra and Query Optimization**

Answer the following questions using queries in relational algebra. For this question, we'll use the following schema, which was described in Question 3.

- Flights (flno, from, to, distance, departs, arrives, price)
- Aircraft (aid, aname, cruisingrange)
- Certified (eid, aid)
- Employees (eid, ename, salary)

a) (3 marks) Find the names of pilots who can operate planes with a range greater than 3,000 miles but are not certified on any aircraft with aname = "Boeing".

b) (2 marks) Consider the following query:

SELECT E.ename

FROM Aircraft A, Certified C, Employees E

WHERE A.aid = C.aid AND E.eid = C.eid AND

Cruisingrange > 5000 AND salary > 100,000

Draw the initial query tree for this query and explain briefly why this plan is considered inefficient.

c) (4 marks) Transform the initial query into an equivalent final query tree that is efficient to execute.

END OF EXAMINATION