HERIOT WATT UNIVERSITY

SCHOOL OF MATHEMATICAL AND COMPUTER SCIENCES

Department of Computer Science

F28DM

Database Management Systems

Semester 2 2016/17

Duration: Two Hours

Answer **THREE** out of four questions

Q1

(a) Define the following terms and give an example of each:

(i) Entity

(1 mark)

(ii) Attribute

(1 mark)

(iii) Relationship

(1 mark)

(b) Draw an Entity Relationship diagram for the following scenario. Identify suitable keys and data types.
Information should be stored about books and their authors. A book will have a title, a unique ISBN number, a date of publication, and one or more authors. An author will have a name consisting of a first and last name.

(12 marks)

(c)

(i) What is an index?

(2 marks)

(ii) Why does a DBMS use an index?

(1 mark)

(iii) Why is it unadvisable to index every attribute of a relation?

(2 marks)

Q2

(a) With the aid of a diagram, explain the steps involved in processing a query.

(8 marks)

(b) Draw an optimised relational algebra tree for the following query. State any optimisation heurisitics you have applied.

```
SELECT D.name
FROM   Department AS D, Employee AS E
WHERE  D.deptno = E.empno AND
       E.surname = 'Gray'
```

The employee table contains 1,000 employees that work for 8 departments.

(10 marks)

(c) State how the following index would affect the query plan in part (b).

```
CREATE INDEX emp_names ON
Employee(surname,firstname)
```

(2 marks)

Q3

(a) State and explain three key advantages of a Relational Database Management System.

(6 marks)

(b) Write an SQL query over the following schema to return the product name and the total number of sales of the product.

```
Product(PID, name, description, cost)
Sales(CID, PID, quantity)
Customer(CID, firstname, surname, address)
```

(5 marks)

(c)

(i) Define the terms online analytical processing and online transaction processing.

(4 marks)

(ii) What is a data warehouse?

(2 marks)

(iii) What operations are performed to populate a data warehouse?

(3 marks)

Q4

(a) Describe how data is stored in a CSV file.

(3 marks)

(b)

(i) Identify the problems with the following XML file and correct them

```
<?xml version="1.0" encoding="UTF-8"?>
<movie>
  <id>1</ID>
  <title>O Brother, Where Art Thou?</title>
  <release-year>2000</release-year>
  <directors>
    <director>Ethan Coen</director>
    <director>Joel Coen</director>
  </directors>
</movie>
<movie>
  <id>2</id>
  <title>Ghostbusters</title>
  <directors>
    <director>Paul Feig</director>
  </directors>
</movie>
```

(4 marks)

(ii) Using your corrected XML file from part (i) write an XPath query to find the names of all movies.

(1 mark)

(iii) Using your corrected XML file from part (i) write an XPath query to find the number of directors.

(2 marks)

(iv) What is meant by XML having a semi-structured data model?

(2 marks)

(c)

(i) State the main motivation behind the NoSQL movement and how this is achieved.

(2 marks)

(ii) Identify three features of a relational database management system that are not guaranteed by a NoSQL system. For each one explain why it is not maintained.

(6 marks)

END OF PAPER