UNIVERSITY OF DUBLIN
TRINITY COLLEGE

Faculty of Engineering, Mathematics and Science

School of Computer Science & Statistics

BA (Mod) Business & Computing **Trinity Term 2013**

Integrated Computer Science Programme

Information Management 1 (CS2041)

Thursday 9th May 2013 Examination Hall 09:30-11:00

Prof. Declan O'Sullivan

Instructions to Candidates:

Answer **two** questions
All questions carry equal marks.
Answer each question in a separate answer book.
Hand up complete question paper.

Materials permitted for this examination:

Calculator

SOME QUESTIONS REFER TO THE FOLLOWING FIGURE A:

```
<?xml version="1.0" encoding="ISO-8859-1"?>
<news>
<news_item timestamp='1'>
  <title> Gorilla Corporation acquires YouNameItWeIntegrateIt.com </title>
  <content>
    <par> Today, Gorilla Corporation announced that it will purchase
YouNameItWeIntegrateIt.com. The shares of YouNameItWeIntegrateIt.com dropped
$3.00 as a result of this announcement. </par>
    <par> As a result of this acquisition, the CEO of YouNameItWeIntegrateIt.com Bill
Smarts resigned. He did not announce what he will do next.</par>
    <par>YouNameItWeIntegrateIt.com is a leading systems integrator that enables
<quote>brick and mortar</quote> companies to have a presence on the web. </par>
  </content>
  <date>1-20-2000</date>
  <author>Mark Davis</author>
  <author>John Smith</author>
  <news_agent>News Online</news_agent>
</news_item>
<news_item timestamp='2'>
  <title>Foobar Corporation releases its new line of Foo products today</title>
  <content>
    <par> Foobar Corporation releases the 20.9 version of its Foo products. </par>
    <par> The President of Foobar Corporation announced that they were proud to
release 20.9 version of Foo products
    </par>
    <figure>
      <title>Presidents of Foobar Corporation and TheAppCompany
      Inc. Shake Hands</title> <image source='handshake.jpg'/>
    </figure>
  </content>
  <date>1-20-2000</date>
  <news_agent>Foobar Corporation</news_agent>
</news_item>
<news_item >
<title>Foobar Corporation is suing Gorilla Corporation for patent infringement </title>
  <content>
    <par> In surprising developments today, Foobar Corporation announced that it is
suing Gorilla Corporation for patent infringement. </par>
    <par>The tension between Foobar and Gorilla Corporations has been increasing
ever since the Gorilla Corporation acquired more than 40 engineers who have left
Foobar Corporation, TheAppCompany Inc. and YouNameItWeIntegrateIt.com over the
past 3 months. </par>
  </content>
  <date>1-20-2000</date></news_item></news>
```

Question 1

a) Define what it is meant by a "NoSQL" approach with respect to the ACID properties of relational data storage systems.

(4 Marks)

b) With the help of diagrams describe how the following different media work to store data: magnetic devices; optical devices; integrated circuits. In addition compare and contrast the advantages and disadvantages of each media in terms of cost and reliability. Finally describe how the file system is used to provide the link between an application or user with data stored on persistent storage in general.

(8 Marks)

c) A file has r=20000 STUDENT records of fixed length. Each record has the following fields: NAME (30bytes), SSN (9bytes), ADDRESS (40bytes), PHONE (9bytes), BIRTHDATE (8bytes), SEX (1byte), MAJORDEPTCODE (4bytes), MINORDEPTCODE (4bytes), CLASSCODE (4 bytes) and DEGREEPROGRAM (3bytes). An additional byte is used as a deletion marker. Block size B = 512 bytes; Block Transfer Time = 1 msec; Rotational Delay = 12.5 msec; Seek time = 30 msec.

i) Calculate the record size R in bytes

ii) Calculate the blocking factor bfr and the number of blocks b assuming an unspanned organisation

iii) Calculate the average time it takes to find a record by doing a linear search on the file if (i) the file blocks are stored contiguously and double buffering is used and (ii) the file blocks are not stored contiguously.

iv) Assume the file is ordered by SSN; calculate the time it takes to search for a record given its SSN value by doing a binary search

(13 Marks)

Question 2

a) Why is it useful to have XML Schema or XML DTD to describe an XML document? Contrast 3 situations where XML Schema and XML DTD differ in their ability to express things?

(4 Marks)

b) Describe in detail the notation that is used in DTDs. Use examples to illustrate your descriptions. In addition, use DTD notation to fully describe the XML document shown in Figure A. Provide explanations for your design decisions.

(8 Marks)

c) Describe what is meant when describing XQuery as consisting of "FLWOR" expressions. In addition, define and explain XQuery statements for each of the following queries posed over the document in Figure A. **Also show expected results and explain your design decisions.**

i) Return any news_item elements that have a timestamp attribute
ii) Return the second <par> element for each news_item
iii) Return in one single element called <authors>, the name of each author mentioned in the document, separated by a semi-colon ";", without the individual <author> tags.

(13 Marks)

Question 3

a) Explain the difference between a B-Tree, B+Tree and B*Tree.

(4 Marks)

b) Explain how indexing can lead to performance gain in information processing. Describe in detail at least 2 indexing approaches, using diagrams to illustrate how they work.

(8 Marks)

c) Given the B-Tree in Figure B (with m=5), show the growth of the B-Tree as you add the keys in the following order
19, 21 22, 24, 25, 28, 5, 33, 38, 16, 45, 56, 26, 58, 70, 73

i. Show the tree at each stage and provide an explanation for any splits
ii. What is the average number of node splits per new key inserted?
iii. What is the average search time?

(13 Marks)

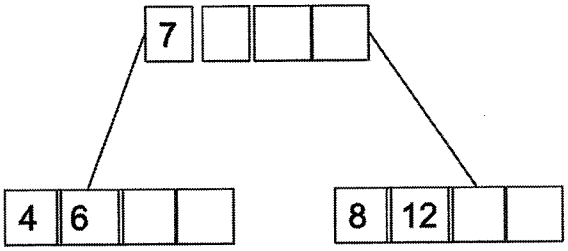

FIGURE B

Question 4

Although UML (Unified Modeling Language) was originally designed to support software development, it is in fact equally useful in supporting Information Modelling. Discuss.

Include at least in your answer: explain the types of UML diagrams that are useful in Information Modelling and what they are useful for; use a simple scenario and example diagrams to illustrate the usage of the diagrams and discuss the advantages/disadvantages of each diagram for modelling; explain how a UML class diagram can be transformed into an XML DTD or XML Schema representation, using examples from your simple scenarios to illustrate.

(25 Marks)