

UNIVERSITY OF BOLTON

**GAMES COMPUTING AND
CREATIVE TECHNOLOGIES**

BSc (Hons) COMPUTING

SEMESTER 1 EXAMINATION 2008/2009

ADVANCED DATABASE SYSTEMS

MODULE NO: CST3003

Date: Thursday 22nd January 2009

Time: 6:00pm – 8.00 pm

INSTRUCTIONS TO CANDIDATES:

There are SIX questions

Answer FOUR questions

Games Computing & Creative Technologies
BSc (Hons) Computing
Semester 1 Examination 2008/2009
Advanced Database Systems
Module No: CST3003

Question 1

- (a) Discuss two advantages and two disadvantages of distributed databases compared with centralised databases. [8 marks]
- (b) Briefly discuss 4 of Date's rules for Distributed Database Management systems, (DDBMSs), and outline compromises that are often made, with respect to Date's rules, when designing a DDBMS. [12 marks]
- (c) As part of a DDBMS, stock data are stored in local databases regarding local stock items only. State the type of fragmentation that has been used within this DDBMS, and explain why such fragmentation is useful. [5 marks]

Please turn the page

Question 2

- a) A decision tree can be used to analyse data as part of a data mining operation. Specify the type of data that is most suitable when using decision trees. With the aid of an example, explain how unsuitable data for a decision tree could be appropriately transformed.

[6 marks]

The table below is to be used to create a decision tree to establish the main factors for mortgage defaulters for a bank.

Number	House style	Deposit	Mortgage type	Defaulted?
1	Detached	Small	Buy-to-let	No
2	Detached	Small	Residential	No
3	Terraced	Small	Buy-to-let	Yes
4	Semi-detached	Small	Buy-to-let	Yes
5	Semi-detached	Large	Buy-to-let	Yes
6	Semi-detached	Large	Residential	No
7	Terraced	Large	Residential	Yes
8	Detached	Small	Buy-to-let	No
9	Detached	Large	Buy-to-let	Yes
10	Semi-detached	Large	Buy-to-let	Yes
11	Detached	Large	Residential	Yes
12	Terraced	Small	Residential	Yes
13	Terraced	Large	Buy-to-let	Yes
14	Semi-detached	Small	Residential	No

- (b) Using the table above and the formulae and $|\log_2|$ table contained in Appendix A, calculate the attribute with the greatest gain and, based on your calculations, draw the first level of the decision tree.

[19 marks]

Please turn the page

Question 3

- (a) An attendance database for a university has a table *Student* with columns: *student_id*, *name*, *telephone_num*, *email*.

When a student leaves the university, their student record is deleted, and a record is created in a table called *Alumni*.

Alumni holds data: *student_id*, *name*, *email*, *withdraw_date*.

A trigger is defined as follows:

```
CREATE TRIGGER student_add_to_alumni
AFTER DELETE ON Student
FOR EACH ROW
BEGIN
  INSERT INTO Alumni
  VALUES (OLD.student_id, OLD.name, OLD.email, CURRENT_DATE);
END
```

- (i) Describe the circumstances under which the above trigger is invoked, and the effect of a successful invocation. [3 marks]
- (ii) Explain what happens if the triggered statement fails (for any reason). [3 marks]
- (iii) Suppose that this trigger is not defined, but the *Alumni* table has been created for the same purpose.

An SQL procedure, called *student_add_to_alumni_proc*, is to be written to make changes to the student table caused by a student leaving (thus its argument is a value for the student identifier) and produce the same effect as the above trigger.

Outline the SQL code for this procedure (exact SQL syntax is not required). [8 marks]

- (b) Explain the differences in the privileges required for a user needing to delete courses when using a pre-written trigger or a pre-written procedure. [6 marks]
- (c) If a group of users requires access to execute the procedure created in (b), outline the appropriate code to achieve this in Oracle. Note: Exact SQL syntax is not required. [5 marks]

Please turn the page

Games Computing & Creative Technologies
BSc (Hons) Computing
Semester 1 Examination 2008/2009
Advanced Database Systems
Module No: CST3003

Question 4

- (a) Describe two of the so-called 'ACID' properties of a transaction. [4 marks]
- (b) With the aid of suitable transaction examples, describe the 'lost update' problem and the 'inconsistent analysis' problem. [10 marks]
- (c) Explain what it means for the concurrent execution of two transactions A and B to be 'serializable'. [4 marks]
- (d) Explain what it means for a schedule to be 'conflict serializable'. With the aid of an appropriate schedule for two transactions, create a precedence graph for a schedule where it is not conflict serializable. [7 marks]

Please turn the page

Question 5.

- (a) Show on a diagram the main components of a data warehouse model, and explain in detail the *extract, transform and load* processes. [6 marks]
- (b) The Course database case study referred to in the module has a simplified ER diagram and entity model as shown on the next page of this examination.

The directors wish to answer questions relating to their main business, which is selling courses to companies, such as:

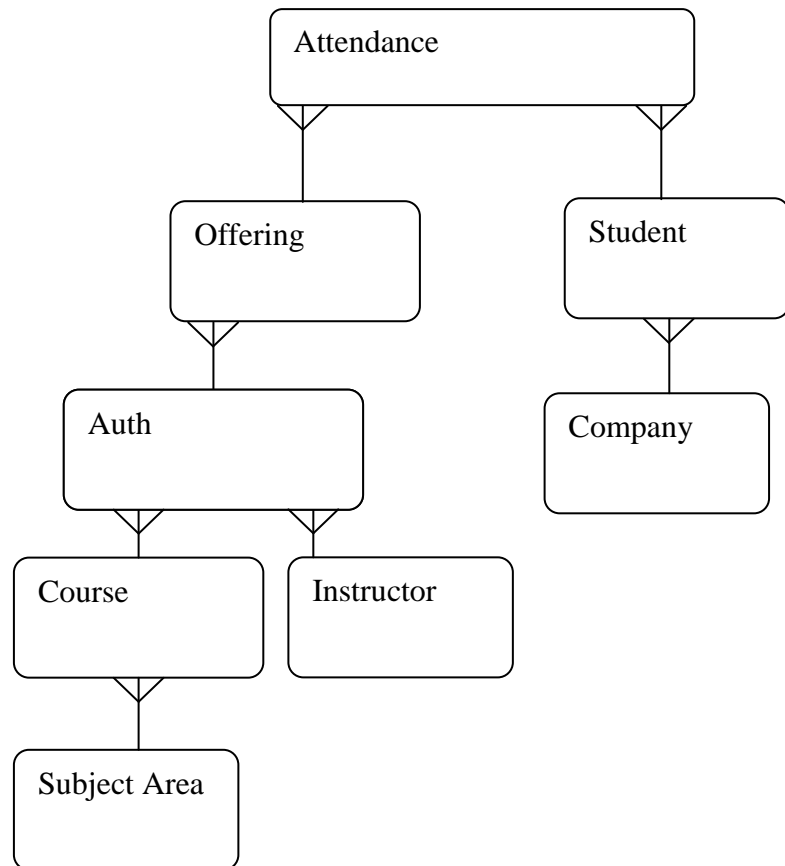
- the courses with the best attendance;
- the most popular courses within the business;
- the companies which have generated most sales.

- (i) For the data warehouse, for such questions to be answered, identify with reasons its facts and dimensions. [7 marks]
- (ii) Draw up an Entity Relationship (star schema) Diagram for the warehouse, and give details of each of the entities and their attributes that would be stored in the warehouse. [8 marks]
- (iii) State four ways in which an operational database is different from a data warehouse. [4 marks]

Question 5 continued over page

Question 5 continued

ER model for Student Database for Question 5.



The entity types in the E-R model are as follows: -

```

Student(student_id, fname, lname, tel_no, gender, date_of_birth, company_id);
Company(company_id, company_name, city, no_employees, revenue, training_b
udget);
Instructor(instructor_id, instructor_name, salary, commission, date_hired);
Course(course_id, title, duration, cost, subject_id);
Auth(instructor_id, course_id, auth_date);
Subject_area(subject_id, subject_name);
Offering(offering_id, start_date, max_no_students, instructor_id, course_id);
Attendance(offering_id, student_id, evaluation, amount_paid);
  
```

Please turn the page

Games Computing & Creative Technologies
BSc (Hons) Computing
Semester 1 Examination 2008/2009
Advanced Database Systems
Module No: CST3003

Question 6.

A major task of a Database Administrator, (DBA), role is the monitoring of a database(s) performance.

Discuss the aspects of a database that would typically be monitored for performance by a DBA, and outline how a DBA could improve upon any perceived performance weaknesses. [25 marks]

END OF QUESTIONS

Please turn the page

Appendix A

For a two-way classification for a decision tree, the relevant formulae are:

$$\text{Entropy, } E = p_+ |\log_2 p_+| + p_- |\log_2 p_-|$$

$$\text{Where: } p_+ = \frac{\text{number}_{\text{positive}}}{\text{number}_{\text{total}}} \text{ and } p_- = \frac{\text{number}_{\text{negative}}}{\text{number}_{\text{total}}}$$

$$\text{and } \text{Gain}(E, A) = E - \sum_{\text{each } v} \frac{n_v}{n_{\text{total}}} E_v$$

Where: A is an attribute and v is an attribute value

| Log₂ | table for numbers from 0 to 0.99 – in 0.01 steps

Number	0	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0		6.644	5.644	5.059	4.644	4.322	4.059	3.837	3.644	3.474
0.1	3.322	3.184	3.059	2.943	2.837	2.737	2.644	2.556	2.474	2.396
0.2	2.322	2.252	2.184	2.120	2.059	2.000	1.943	1.889	1.837	1.786
0.3	1.737	1.690	1.644	1.599	1.556	1.515	1.474	1.434	1.396	1.358
0.4	1.322	1.286	1.252	1.218	1.184	1.152	1.120	1.089	1.059	1.029
0.5	1.000	0.971	0.943	0.916	0.889	0.862	0.837	0.811	0.786	0.761
0.6	0.737	0.713	0.690	0.667	0.644	0.621	0.599	0.578	0.556	0.535
0.7	0.515	0.494	0.474	0.454	0.434	0.415	0.396	0.377	0.358	0.340
0.8	0.322	0.304	0.286	0.269	0.252	0.234	0.218	0.201	0.184	0.168
0.9	0.152	0.136	0.120	0.105	0.089	0.074	0.059	0.044	0.029	0.014

For example, | log₂ | of **0.32** = **1.644**

Look along row 0.3 and select the value at column 0.02

Note: You should round numbers and, if they are between values round up,
 e.g. for **0.465** round to **0.47** to get value **1.089**