

UNIVERSITY OF BOLTON – RAK CAMPUS

**SCHOOL OF BUSINESS AND CREATIVE
TECHNOLOGIES**

BSc (Hons) COMPUTING

SEMESTER 1 EXAMINATION 2009/2010

ADVANCED DATABASE SYSTEMS

MODULE NO: CST3003

Date: Wednesday 20th January 2010

Time: 2:00 – 4.00 pm (2 hours)

INSTRUCTIONS TO CANDIDATES:

Answer FOUR questions

Business & Creative Technologies
BSc (Hons) Computing
Semester 1 Examination 2009/2010
Advanced Database Systems
Module No: CST3003

Question 1

- (a) An advantage of distributed databases, when compared with centralised databases, is said to be their improved performance. Outline the difference between a distributed database and a centralised database, and with the aid of a relevant example, explain why a well-designed distributed database system can perform better than a centralised database system. [8 marks]
- (b) A Distributed Database Management system, (DDBMS), can be designed with 'complete replication' or 'selective replication'. Briefly describe each approach and outline circumstances where 'complete replication' would be more appropriate for a business than the 'selective replication' approach. [10 marks]
- (c) As part of a DDBMS, all salary data columns are stored in one database, whilst partial salary data columns are stored in other databases. State the type of fragmentation that has been used within this DDBMS, explain how this might be achieved, and explain why such fragmentation is useful. [7 marks]

Please turn the page

Business & Creative Technologies
 BSc (Hons) Computing
 Semester 1 Examination 2009/2010
 Advanced Database Systems
 Module No: CST3003

Question 2

- a) A decision tree can be used to analyse data as part of a data mining operation. Briefly outline what a decision tree is and outline the main advantages of their use within a business environment. [6 marks]

The table below is to be used to create a decision tree to establish the main factors for a viral marketing campaign of a social networking site.

Number	Age category	Number of friends	Gender	Responded?
1	Young	High	Male	No
2	Young	High	Female	No
3	Mature	High	Male	Yes
4	Adult	High	Male	No
5	Adult	Low	Male	No
6	Adult	Low	Female	No
7	Mature	Low	Female	No
8	Young	High	Male	No
9	Young	Low	Male	Yes
10	Adult	Low	Male	Yes
11	Young	Low	Female	Yes
12	Mature	High	Female	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

Business & Creative Technologies
BSc (Hons) Computing
Semester 1 Examination 2009/2010
Advanced Database Systems
Module No: CST3003

Question 3

- (a) Explain the options available, in Oracle, for where or how user-defined PL/SQL procedures are stored. [6 marks]
- (b) An insurance company uses an Oracle database which has a table *Quotation* with columns:
quoteNum, custNum, quoteItem, quoteDate, quoteDetail, quoteValue.

When a customer rejects the offer from an insurance quotation, their quotation record is deleted and a record is created in a table called *ReQuote*.

ReQuote holds data: custNumber, quoteItem, quoteDate.

A trigger is defined as follows:

```
CREATE TRIGGER customer_add_to_requote
AFTER DELETE ON Quotation
FOR EACH ROW
BEGIN
    INSERT INTO ReQuote
    VALUES (OLD.custNumber, OLD.quoteItem, OLD.quoteDate);
END
```

- (i) Describe the circumstances under which the above trigger is invoked, and the effect of a successful invocation. [3 marks]
- (iii) Suppose that this trigger is not defined, but the *ReQuote* table has been created for the same purpose.

An SQL procedure, called *customer_add_to_requote_proc*, is to be written to make changes to the *Quotation* table caused by a customer rejecting a quote (thus its argument is a value for the quotation number) and produce the same effect as the above trigger.

Outline the SQL code for this procedure (exact SQL syntax is not required). [10 marks]

- (c) Explain the differences in the privileges required for a user needing to delete quotations when using a pre-written trigger or a pre-written procedure. [6 marks]

Please turn the page

Business & Creative Technologies
BSc (Hons) Computing
Semester 1 Examination 2009/2010
Advanced Database Systems
Module No: CST3003

Question 4

- (a) State each 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 'uncommitted dependency' problem. [10 marks]
- (c) Explain what it means for a schedule to be 'conflict serializable' and briefly outline how such schedules can be identified. [4 marks]
- (d) Transaction concurrency is often controlled using locking mechanisms. In the context of locking, and using an appropriate example, explain how deadlock may occur. Outline three techniques for resolving deadlock. [7 marks]

Question 5.

- (a) Describe three potential benefits of Data Warehousing. [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:

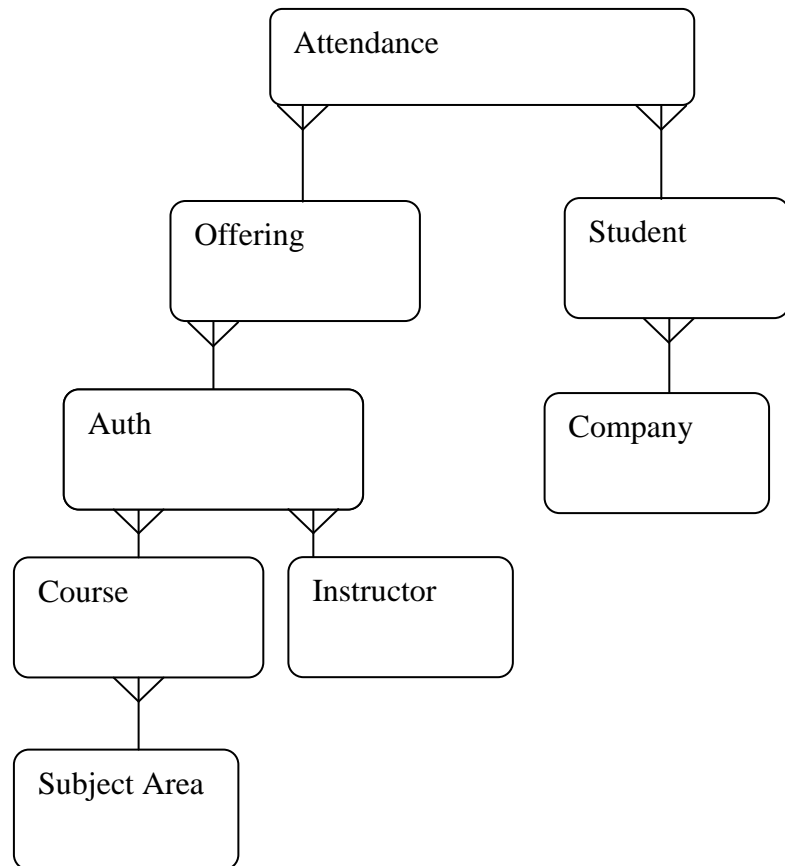
- the least popular subject areas;
 - the most popular courses within the business;
 - the companies that generate most repeat business.
- (i) For the data warehouse, for such questions to be answered, identify with reasons its facts and dimensions. [7 marks]
- (ii) Draw up an Entity Relationship (star 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

Business & Creative Technologies
 BSc (Hons) Computing
 Semester 1 Examination 2009/2010
 Advanced Database Systems
 Module No: CST3003

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_budget);
 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

Business & Creative Technologies
BSc (Hons) Computing
Semester 1 Examination 2009/2010
Advanced Database Systems
Module No: CST3003

Question 6.

A major task of a Database Administrator, (DBA), role is the monitoring and maintenance of security.

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

END OF QUESTIONS

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**