TECHNOLOGICAL UNIVERSITY DUBLIN
CITY CAMPUS

––––––––––

TU856/DT228 BSc in Computer Science
TU857/DT211C BSc in Computer Science (Infrastructure)
TU858/DT282 BSc in Computer Science (International)
Year 4

––––––––––

SEMESTER 1 EXAMINATIONS 2021/2022

––––––––––

Advanced Databases

Internal Examiners:
Deirdre Lawless
Giulia Vilone
Dr. Paul Doyle

External Examiners:
Sanita Tifentale – TU856/DT228
Pauline Martin – TU857/DT211C
Pamela O'Brien – TU858/DT282

Instructions

Answer any FOUR questions.

All questions carry equal marks.

1. **(a)** Explain the key differences between normalization and denormalization.

    (5 marks)

    (b) Explain *FOUR (4)* possible benefits of denormalization.

    (4 marks)

    (c) A gardening service company (GSC) undertakes a range of tasks for clients and supplies the products used during tasks e.g., weedkiller, plants, seeds etc.). Consider the following ER Schema (primary keys are underlined, foreign keys are indicated by the suffix (FK)):

    ```
    Client(clientId, clientName, clientAddress, phone, email,
    contact_person)
    GardenStaff(staffId, first_name, last_name)
    Garden_task(taskId, clientId (FK), startTime, endTime, staffId
    (FK), success_or_fail)
    Product(productId, productName, unit, pricePerUnit)
    Product_supplied(psId, productid (FK), taskId (FK), units)
    Product_offered(poId, productid (FK), taskId (FK), units)
    Product_writtenoff(woId, woDate, productId (FK), units, wo_reason)
    Products_on_order(orderId, productId (FK), suppId (FK), units)
    Supplier(suppId, suppName, suppAddress, phone, email,
    contact_person)
    ```

    GSC can undertake many tasks for a client. Each client nominates a contact person, but this may change over time. For each task, GSC recommends a number of products to the client (product_offered) and records the products the client actually selects (product_supplied). When a task is completed, the client indicates whether they were satisfied with the service or not. GSC maintains a store of products they use and regularly needs to dispose of products that are no longer usable. This represents a loss to GSC and details are recorded (product_writtenoff). Products need to be replaced regularly. Details of any orders placed with suppliers are recorded (products_on_order). GSC would like to be able to generate reports on their annual performance.

    (i) Provide an example of how denormalization could be applied to this schema to achieve each of the possible benefits you identified in part (b).

    (4 x 3 marks)

    (ii) What are the possible negative implications of the examples you provided in your answer to part (c) (i)?

    (4 marks)

2. (a) Other than auditing, explain *FOUR (4)* possible uses of triggers in a relational DBMS.

(4 marks)

(b) Consider the following trigger and data:

```
CREATE OR REPLACE TRIGGER credit_limit_setup
BEFORE update ON customers
FOR EACH ROW
BEGIN
    IF upper(:NEW.vip) ='N' and (:NEW.creditlimit > :OLD.creditlimit*2) THEN
        RAISE_APPLICATION_ERROR(-20001, 'Warning: non-VIP credit limit cannot
        be increased by a factor greater than 2 ');
    END IF;
    IF upper(:NEW.vip) ='Y' and (:NEW.creditlimit > :OLD.creditlimit*3) THEN
        RAISE_APPLICATION_ERROR(-20002, 'Warning: VIP credit limit cannot be
        increased by a factor greater than 3');
    END IF;
END;
```

Customers			
cust_id	custname	creditlimit	VIP
1	cust1	662	Y
2	cust2	703	N
8	cust8	651	N
10	cust10	815	Y

(i) Explain the purpose of a trigger statement. Identify the trigger statement for the trigger provided.

(2 marks)

(ii) Explain the purpose of a trigger action. Identify the trigger action for the trigger provided.

(3 marks)

(iii) Explain what would happen if each of the SQL statements in the table below were executed against the data shown above. Justify your answers.

SQL Statements
update customers set creditlimit=1500 where cust_id=2;
update customers set creditlimit=1600 where cust_id=10;
update customers set creditlimit=2000 where cust_id=11;

(3 x 2 marks)

(c) Write a trigger to ensure that whenever a record is inserted into a sales table, a record is created in a `sales_log table (sales_id, insertedby, insertime)`. The identifier of the sales record inserted, the username that executed the SQL to create sales record, and the date and time the sales_log record was created are recorded. Note: You can retrieve the username from dual e.g. SELECT user FROM dual; SYSTIMESTAMP holds current date and time.

(10 marks)

3. (a) (i) Explain how each of the following works in Oracle's DBMS and when they would be used:

    a. Full index scan

    b. Fast full index scan

    c. Index range scan

(3 x 2 marks)

(ii) Consider the following scenarios. For each, explain which type of scan listed in part (i) will be used by the Oracle DBMS and why:

- Suppose the last_name column has a not null constraint and the last name and clientFee are a composite key in an index and the query SELECT last_name, clientFee FROM clients; is executed.

- Suppose that department_id, last_name, and salary are a composite key in an index and the query SELECT department_id, last_name, clientFee FROM clients WHERE clientFee > 5000 ORDER BY department_id, last_name; is executed.

- Suppose the last_name column is indexed, there are clients whose last name start with A, and the query SELECT last_name FROM clients WHERE last_name like 'A%'; is executed.

(3 x 3 marks)

(b) (i) Explain *THREE (3)* types of index that can be implemented in a MongoDB.

(3 x 2 marks)

(ii) Suppose we have a collection student containing the following type of document:

```
{
  "_id": 1,
  "person": { name: "Jane", surname: "Byrne" },
  "age": 20,
  "city": "Waterford"
}
```

Provide the code to create an index which would improve queries on name, surname and city. What type of index is this?

(4 marks)

4. Consider the case of a real estate agency whose database is composed by the following tables:

```
OWNER (IDOwner, Name, Surname, Address, City, Phone)
PROPERTY (IDProperty, IDOwner, PropertyCategory, Area,
City, Province, Rooms, Bedrooms, Garage, Meters)
CUSTOMER (IDCust, Name, Surname, Budget,
Address, City, Phone)
AGENT (IDAgent, Name, Surname, Office, Address, City, Phone)
AGENDA (IDAgent, Data, Hour, IDEstate, ClientName)
VISIT (IDProperty,IDAgent, IDCust, Date, Duration)
SALE (IDProperty,IDAgent, IDCust,Date, AgreedPrice, Status)
RENT (IDProperty,IDAgent, IDCust, Date,Price, Status, Time)
```

Your goal is to define a data warehouse to provide a supervisor with an overview of the situation. The supervisor must have a global view of the business, in terms of the properties the agency deals with and of the agents' work. The supervisor is very interested in the sales trend and controlling agents' time.

Here are some examples of queries it should be possible to execute:

a. How many customers have visited properties of at least 3 different categories?

b. What is the average duration of visits per property category?

c. Who has paid the highest price among the customers that have viewed properties of at least 3 different categories?

d. Who has bought a flat for the highest price each month?

e. What kind of property sold for the highest price in each city and month?

(a) **(i)** Provide a star schema which will achieve the goals outlined.

(15 marks)

(ii) Explain how your schema will facilitate the queries listed a to e above.

(5 marks)

(b) Write the SQL needed to allow the supervisor to identify the average duration of visits per property category.

(5 marks)

5. You are required to design a database to store information about army soldiers and their chain of command. Each soldier is described by a soldier_id, a soldier name and a salary. There is a relationship among the soldiers: soldier *a* is linked to soldier *b* if a is *b*'s captain. The date soldier *a* became *b*'s captain is also stored.

(a) Provide a relational schema to show how to store soldiers and their relationship. Provide tables, fields and show primary and foreign keys.

(4 marks)

(b) Write an SQL query to get all the list of soldiers directly commanded by the soldier with id=3.

(4 marks)

(c) Write an SQL query to get the list of soldiers commanded by soldier with id=3 directly <u>or</u> commanded by soldiers managed by soldier with id=3.

(5 marks)

(d) Provide a json structure to store the same information provided in the relational model into a *mongodb* collection.

(4 marks)

(e) Compare the two data models: which one is easier to query?
Is standard SQL a sustainable way to perform such queries?

(4 marks)

(f) Show how the same information would be stored in a graph database. Why could this be considered a better solution than using MongoDB?

(4 marks)