Full Time and ACCS

Semester Five : January 2021

Big Database Technologies

Internal Examiners
Mr. Seán Mc Hugh
Mr. Phillip Fitzpatrick

External Examiners
Mr. Thomas Davis

Day   **Wednesday**
Date   **13th January 2020**
Time   **09.30-11.30**

Instructions to Candidates

Please answer any three out of four questions on the paper. Note Question 1 carries 34 marks. All other questrions carry 33 marks.

Please start each question on a new page

QUESTION 1 (34 Marks)

a) The Object Relational extension to the SQL standard provides an object-oriented approach to storing data in the relational database. With the aid of an example, explain how data can be stored using this approach.

Discuss the merits of this approach compared with the Relational Approach. Is there a disadvantage to this approach? Provide examples to demonstrate your answer.

(16 Marks)

b) Tallaght University Hospital provide heart monitors to patients for a week's duration. The Heart monitor device stores heart readings every 10 seconds. When the monitor is returned to the hospital, data is uploaded to their Relational database tables below. It has been decided to rationalise the tables.

The Heart Reading table is to be removed and its data is to be stored in the Patient Heart Monitor Device table. Discuss the object relational options available to you outlining their characteristics. Which one is the most suitable for this situation? Explain your reasoning. Provide sample code to support your answer.

(18 Marks)
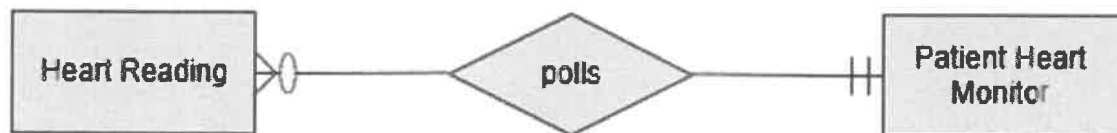
HeartReading(timestamp, sensorType, heartRate)

PatientHeartMonitor(patient id, deviceId, deviceName, deviceModel)

QUESTION 2 (33 Marks)

You have been hired as a database specialist because of your experience on critical disparate database platforms in industry.

a) Provide a business case for Relational Databases. What is the business case for NoSQL databases? Give examples to support your answers. Can these database technologies be used in the same application?

(16 Marks)

b) It is planned to model the Discussion Forum below using the MongoDB database. Discuss the design options available to you. Which is the optimal solution and why? State your assumptions and explain your reasoning.

Forum Discussion(id, Forum_Category, Title, Text_Body, Poster_Name, Date)
Posts(User_id Poster_name, Date, Post_Body)

(17 Marks)

QUESTION 3 (33 Marks)

a) Compare and contrast Document Databases with Column Family Stores regarding the structure of how the data is stored and managed. Provide relevant diagrams to demonstrate your answer.

(9 marks)

b) Facebook, using neural network techniques, classify billions of user images and user posts with the aim to identify people, predict their current emotions and target them with personalised advertisements. In this context, explain the characteristics of Big Data.

(10 Marks)

c) Riak's Key-Value store architecture uses a logical ring. Discuss the structure of the ring. As part of your answer, highlight the significance of Replication, Consistent Hashing and the Hinted Hand-Off in the operation of the ring. Why are they important?

**(14 Marks)**

QUESTION 4 (33 Marks)

Data warehouse data can be described as subject-oriented, integrated, time-variant, and a non-volatile collection of data in support of management's decision-making process (Inmon, 1993).

a) Critically appraise the characteristics of data in a data warehouse as defined in the above description. Contrast them with the characteristics of data in an OLTP system. Provide examples to support your answer.

(12 Marks)

b) Using the dimensional model below, explain its main characteristics. Discuss the main reasons for using this model. Are there any drawbacks?

(11 Marks)
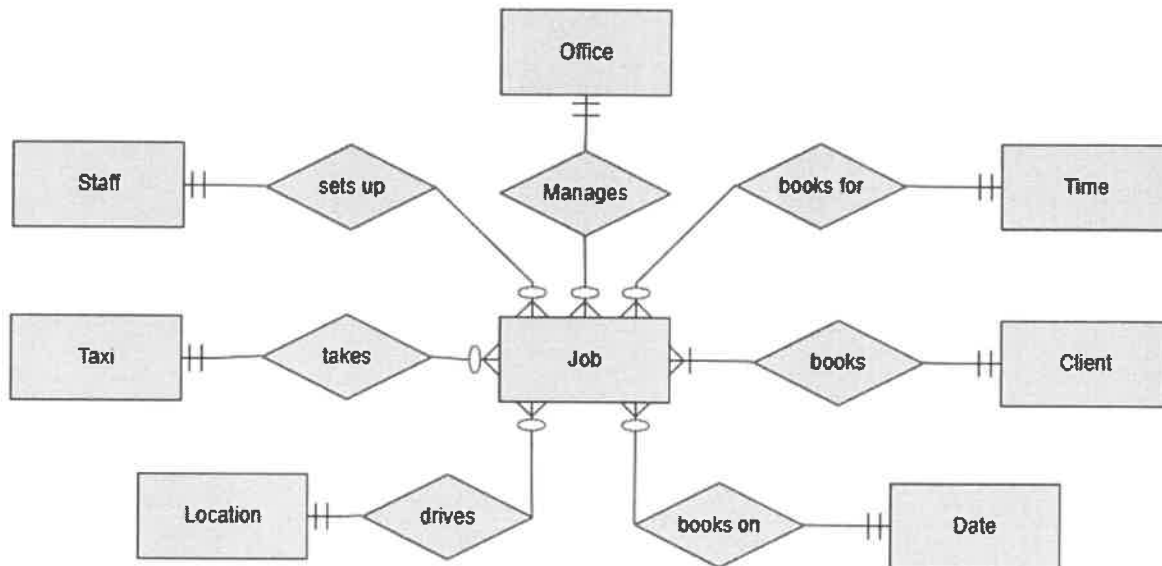
Office (officeID(PK), officeno, state, city, postcode, managerName)
Time (timeID(PK), 24hourclock, am/pmIndicator)
Client (clientID(PK), clientNo, fullName, street, city, postcode)
Date (dateID(PK), dayofWeek, week, month, quarter, season, year)
Location (locationID(PK), postcode, area, town, city, region)
Taxi (vehRegID(PK), vehRegNo, model, make, colour, capacity)
Job(jobID, mileage,totalCharge)

c) The following SQL represents a common query in the Data Warehouse

```
SELECT d.quarter, t.vehRegNo, l.area, SUM(j.charge)
FROM job j, taxi t, date d, location l
WHERE t.vehRegID = j.vehRegID
AND    d.dateID = j.dateID
AND    l.locationID = j.locationID
AND    l.city ='Cork'
GROUP BY d.quarter, t.vehRegNo, l.area
```

Discuss any techniques which are particularly relevant to a Data Warehouse environment that would improve the performance of this query. Which technique would you recommend and why?

(10 Marks)