Tutor-marked assignment TMA 02

This item contains selected online content. It is for use alongside, not as a replacement for the module website, which is the primary study format and contains activities and resources that cannot be replicated in the printed versions.

Contents

Introduction 2
- Preparing your TMA 2
- Learning outcomes 4

Part 1: Quiz results (5 marks) 5
- Question 1 (5 marks) 5

Part 2: Database management (20 marks) 6
- Question 2 (20 marks) 6

TMA02 Scenario overview 7

Part 3: Data modelling (55 marks) 9
- Question 3 (35 marks) 9
- Question 4 (20 marks) 11

Part 4: SQL (120 marks) 12
- Question 5 (20 marks) 12
- Question 6 (65 marks) 14
- Question 7 (35 marks) 16

Introduction

This tutor-marked assignment (TM254 TMA 02) must be submitted by 12 noon (UK local time) on the cut-off date **12 March 2020**.

This module requires all assignments to be submitted electronically. To submit an assignment, please follow the link(s) from your StudentHome page to the online TMA service.

If you foresee any difficulty with submitting your assignment on time, you should contact your tutor well in advance of the cut-off date.

For further information about policy, procedure and general submission of assignments please refer to the Assessment Handbook, which can also be accessed via your StudentHome page.

This assignment accounts for 30 per cent of the assessment score for this module. It has seven questions, for which a total of 200 marks are available. TMA02 is marked out of 200 points and is worth double the marks of TMA01 and TMA03. Questions do not all have the same number of marks; the marks allocated to each question or part question are indicated.

Submission document

You are required to produce a single word-processed document containing all your work, which we refer to as your *solution document*.

Make sure that you create and submit your answers in a file with an acceptable file format. Acceptable file formats are those with extensions of .doc, .rtf or .docx.

Preparing your TMA

Installing the drone database

For question 6 in Part 4 of this TMA you will write SQL queries to satisfy data retrieval requests. We've created a sample database against which you can run your queries and generate answers that you can paste into your TMA solution document.

To install the drone tables, you should:

1. Locate the file drone.txt in the Assessment area of the module website; then download this file, placing it in a folder so that it can be evaluated by the Postgresql system using the \i command in psql.
2. Connect to Postgresql using psql: use the default postgres database, default port, and default postgres user with the appropriate password (TM254 if you followed the installation suggestion in the Software Guide).
3. Use the `\i drone.txt` command in psql to load and run the SQL code, this will create the tables and data for the drone sample database. (Note you may need to give the full path to the location where you placed the drone.txt file: `\i 'C:\Users\....drone.txt'`)

4. To test the installation you should run the query `SELECT * FROM drone;` in psql. The resulting table should have 6 rows as shown below.

```
manufacturer |   serial     |  type    | purchase_date
-------------+--------------+---------+---------------
 RotorDyne     | RD142566243 | RX12     | 2019-06-21
 RotorDyne     | RD142562324 | RX12     | 2019-06-21
 CyberDyne     | Model 101   | T-800    | 1984-05-10
 Tyrell        | 99544       | Nexus-6  | 2019-08-24
 FlyteSpeed    | 1442522     | Evoc1    | 2019-09-20
 MyFly         | BTTF2       | H-Board  | 2019-10-21
(6 rows)
```

You only need to evaluate the drone.txt file's content once, the tables will remain in the postgres database if you leave then reconnect using psql.

When working with Question 6 you may want to download the document drone_schema (in the assessment resources) as a handy reminder of the tables and foreign keys in the sample database.

Word limits and layout

There are questions in TMA 02 with specified word limits. Exceeding a word limit by more than 10% will lead to a deduction of 2 marks for that question or part question. If a question or part question has a word limit, state how many words you used. You should aim to write clearly and concisely in English using appropriate terminology for the subject material.

Some answers, for example those with SQL table output, require you to copy and paste text from other sources. Please paste this text into the body of your solution document along with the rest of the answer for the relevant question. Do not place any pasted text separate from your main answer, and do not put it in an appendix or additional document(s).

Remember to ensure that you include a header or footer on each page of your solution document that includes your Open University identifier, your name and the page number; this helps if your tutor needs to print your solutions while marking.

Referencing

Where an answer draws on the module parts or the set books you should cite the source, for example (TM254 Block 1 Part 5) or (Agutter, Chapter 1, Stakeholders) or Hughes,

Chapter 1, Section 2.2). You do not need to give a corresponding reference at the end of your solution document.

Any other sources you use should be cited and fully referenced using the OU Harvard style provided on the Library's Referencing and plagiarism page.

Accessibility and software use issues

If you require additional support for software or with accessing the assessment content – for example with diagram interpretation, drawing package use and suchlike – then please inform your tutor at the earliest opportunity.

Similarly, if you are unable to meet any of the submission requirements then please contact your tutor at the earliest opportunity. Your tutor may not be able to access your solution document if you choose an alternative format for submission; with their advice, you may be able to agree an acceptable alternative. The Technical Help forum is also a good place to discuss issues with the process of submission – but please do not discuss the content of TMAs in the module forums.

Learning outcomes

This assignment assesses learning connected with the following outcomes.

Knowledge and understanding

- KU2 Demonstrate knowledge and understanding of capturing and understanding the requirements of stakeholders and be able to satisfy some of those requirements with a database solution.

Cognitive skills

- CS3 Understand the importance of database administration.

Key skills

- KS2 Communicate effectively about service and project management, requirements and database design.

Practical and professional skills

- PPS3 Use a standard database environment to define and implement a database solution.

Part 1: Quiz results (5 marks)

This question asks you to summarise your engagement with the Block 2 quiz questions. You can complete this question after you have attempted the Block 2 quiz.

Note that the marks you will receive for this question are not based on your final quiz score, but on your reflections on engaging with the questions. You are asked to supply a screenshot of the progress you made through the quiz, but you will not be penalised for any questions you got wrong.

Question 1 (5 marks)

a. Take a screenshot of the question navigation panel from the Block 2 quiz, showing the full extent of your engagement with the questions. Your screenshot should look similar to Figure 1. Paste the screenshot into your solution document as your answer to this part of the question.

Figure 1 A screenshot of the question navigation panel (yours may look different from the example shown)

(2 marks)

b. *You should write no more than **90 words** for this part of the question.*

Write a few sentences in which you tell your tutor about your experience of completing the quiz. For instance, you may have found some Block 2 quiz questions particularly interesting or challenging. In that case, tell your tutor why you found these questions challenging or interesting. Think of this as an opportunity to communicate with your tutor about your experience of studying Block 2.

(3 marks)

Part 2: Database management (20 marks)

This question can be answered after you have completed your study of Block 2 Part 3 and any related activities within that part.

It is intended to allow you to show your understanding of the context of data management and the features found in a modern database management system. In addition, you should be able to show that you understand the importance of data independence and how that relates to the three-schema data architecture.

Question 2 (20 marks)

a. State why data has both cost and value to an organisation.

(2 marks)

b. *You should write no more than **150 words** for this part of the question.*

Briefly explain why a database management system (DBMS) is much more than just a database

(6 marks)

c. An existing relational database, developed in accordance with the three-schema architecture, is supporting several application processes that share data. Some new requirements are identified by the database users that result in new applications being developed, as follows:

   a. A new application process requires a new table in the database and uses none of the existing tables.

   b. A new application process is required that uses data already in use by an existing application process. However, the performance requirements for the new application require the database to use new hardware so that data retrieval is more efficient.

   c. A new application uses tables already in the database that are currently in use by different application processes; no existing application process uses all these tables.

   For each of these application's external schema requirements:

   i. state what changes would be required (if any) to each of the existing schemas in the three-schema architecture, and state why that change is required

   ii. state whether or not the required schema changes would affect any existing application processes and justify your answer.

   (4 marks for each application. 12 marks total)

TMA02 Scenario overview

This scenario is supplied to provide context for the TMA questions in Parts 3 and 4, it will help with interpreting the scenario examples used in questions 3-7.

A company is offering a commercial drone service to clients. The company has a fleet of drones and a range of data capture devices that the drones can carry. The range of services they offer includes still image and video capture of events and locations. For example, weddings, marketing images, sports and entertainment events, building surveying involving mapping and photographing buildings and structures in detail, and site surveying involving area mapping and feature extraction. The company agrees contracts with clients that describe the service events – the company supplies the drones, pilots and observers, undertakes the required tasks for the client at a specified location and then supplies the client with the resulting data files generated during the activity. The data files themselves are generated on high capacity SD cards, and long-term storage is on a large fileserver rented from a cloud service provider.

The company wants to maintain historic records of the activities it undertakes, including the drone used, pilots and observers, maintenance records and data files generated. In addition, they are required to keep flight records for drones flown on company business including training and test events.

At each event, which takes place at a specific location on a stated date, there will have been a drone, a pilot and an observer. The company could be holding several events simultaneously. An event may be a training or test flight, in which case there is no client involved. If an event was for a client then it will have been booked for a single client (no event involves more than one client), and each client might have booked several events. The event may generate multiple data files, each file is specific to that event – the company wants to record each file's metadata (unique filename, file type, date and time recorded, data size and the file's location on the cloud storage) and link it to the event at which the file was generated.

The company's drones undergo three types of maintenance – commissioning (checking new drones and approving them for commercial use), regular and post-incident (repairs following a crash, impact or anomalous behaviour incident). A maintenance session is for one drone and the records show the type of the maintenance, the start and end dates of the maintenance period, the technician who performed the maintenance, and text notes summarising what actions have been taken. They also note the technician who signed off the drone at the end of the maintenance session (who cannot be the same technician who performed the maintenance). A record is made of either the date on which the completed maintenance was checked, or the date on which the drone was written off (no longer to be used for commercial activities) – only one of these can occur.

Note we've made some simplifications for this scenario. The company aren't tracking/recording battery usage with the drones, or which data capture devices were used. The requirement is for historic records, i.e. after the events, which removes the issues that always arise between planning an event (which pilots are available, which pilot has been booked, etc) and the reality of an event (such as cancelling events due to bad weather, substitute pilots replacing the booked pilot, and such like.). The company also simplified the maintenance records by assuming they are complete records when they are entered into this database (so start and end dates and sign-off or write-off dates are all present and inserted at the same time).

PLEASE ENSURE YOU HAVE INSTALLED THE `drone` DATABASE BEFORE YOU BEGIN QUESTION 6.

Part 3: Data modelling (55 marks)

The two questions in this part of the TMA are based around Block 2 Parts 4 and 5. They focus on the different ways that data can be modelled, including conceptual data modelling, entity-relationship diagramming and relational representation.

Question 3 is worth 35 marks and Question 4 is worth 20 marks.

Question 3 can be answered after you have completed your study of Block 2 Part 4 and any related activities within that part. It is intended to allow you to demonstrate your understanding of the purpose of conceptual data modelling and its terminology. You will also demonstrate your ability to produce and interpret entity-relationship notation.

Question 4 can be answered after you have completed your study of Block 2 Part 5 and any related activities within that part. It is designed to allow you to demonstrate your understanding of the purpose of the logical schema, and the use of the relational model to represent degree and optionality from the entity-relationship notation. In addition, the question allows you to demonstrate your ability to interpret and develop a relational representation from a conceptual data model.

Question 3 (35 marks)

a. Read the following description related to the drone scenario:

> The company works with clients who get them to undertake drone flights to capture data recordings at specified locations.
>
> The company keeps information about clients, their company name, the name of the contact for communication with the client, their address, phone number and date of first contact. They give each client a unique identifier, which allows for more than one contact within a large organisation.
>
> Clients enter into a contract that involves work that the company calls an event. The record for an event shows where it takes place, which employees were pilot and observer at the event, the date on which the event took place, and there is a description of what is expected at the event (for example a description of the required data capture (e.g '3D surface mapping of pylon allowing 0.5 cm resolution', 'Ground radar data of 30 metre grid centred on marker post, with corresponding full colour images'…). The company creates a unique event identifier for each event.
>
> The company keeps track of the location at which events have been booked. They record the GPS location (latitude and longitude values, to 5 decimal places which identifies an area roughly 1 metre square) and there's a description of the location so that the specific structure or space is clearly identified. A note of any relevant flight restrictions and/or hazards is also held (e.g. 'within 1 km of active small airfield', 'large expanse of open, deep water, untethered balloon flights launched 3 km NE of site').

List the entity types and relationship types that can be identified from the information given in the description at the start of part a) immediately above. For the entity types, give their attributes and indicate a possible identifier (primary key) for each entity type.

Use the notation `EntityType(Attribute1, Attribute2, Attribute 3 …)` for entity types, their attributes and the identifier; and the notation `EntityType RelationshipType EntityType` for the relationships.

You are *not* required to draw an entity-relationship diagram, nor are you required to identify any degree or participation conditions for the relationship types.

(10 marks)

b. Figure 2 shows instances of entity types and a relationship type (You have seen similar diagrams in Block 2, Part 4).

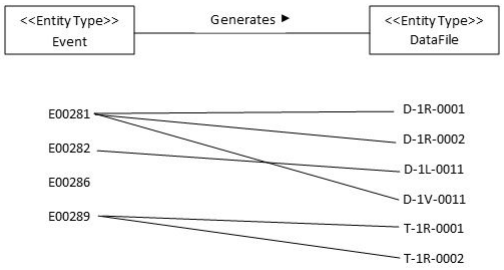

Figure 2 An occurrence diagram

i. How many relationship instances are shown in Figure 2 and how many entity instances, of each type, are involved with these relationship instances?

(3 marks)

ii. Write down sentences, of the form used in Block 2 Part 4, Section 3 (particularly section 3.2.6, Exercise 4.2) that describe the relationship type in Figure 2. Your sentences should capture the degree (multiplicity) and participation conditions (optionality).

(4 marks)

iii. Draw an entity-relationship diagram that captures the information given in your sentences describing Figure 2. (Your diagram must show multiplicity and optionality.)

(2 marks)

c. *You should write no more than **60 words** for this part of the question.*

Briefly describe how the data descriptions you produced in your answers to parts a) and b) above would be used when working with stakeholders to develop a database.

(4 marks)

d. Consider the following description.

> After a new drone is purchased it is commissioned for commercial use at a maintenance session. During its commercial use with the company it may, when necessary, be repaired several times at different maintenance sessions. Eventually the drone is no longer cost effective to repair and is written off at the end of a maintenance session, after which the company no longer use the drone. Each maintenance session involves a single drone for either commissioning, regular maintenance, repair or writing off.

A data analyst has considered the information in the above description and has generated Figure 3 showing the Drone and Maintenance Session entity types and some relationship types that could hold between them – however, the data analyst has not done a good job.

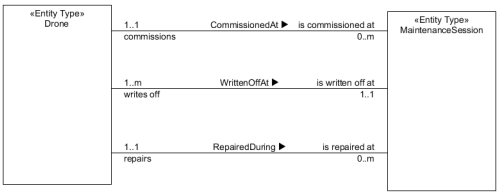

Figure 3 An entity-relationship diagram

For each relationship type in Figure 3, write down the sentences that give a reading to the relationship type. (You have seen sentences of the form required in Block 2 Part 4 Section 3, particularly section 3.2.6, Exercise 4.2).

Do the degree (multiplicity) and participation (optionality) seem appropriate for the relationships named? For each relationship type where you feel they are not appropriate, justify your answer by stating any problem that Figure 3 contains (you do not have to correct the diagram).

(12 marks)

Question 4 (20 marks)

a. Primary keys and foreign keys are important in relational representations.

Briefly explain:

i. what they are, and how they relate to each other.

ii. why foreign keys and the referential integrity rule are important in representing relationships within a database.

iii. why declaring a primary key represents a constraint on the values possible in a relation.

(8 marks)

b. Figure 4 is a fragment of a conceptual data model consisting of an entity-relationship diagram showing entity types Pilot and Event, and the relationship type InChargeOf between them.

Figure 4 A fragment of a conceptual data model

i. Convert Figure 4 into a logical schema representation using the relation-for-relationship representation.

ii. Convert Figure 4 into a logical schema representation using the posted-key representation.

You can use either diagrams or text representations for the logical schema.

(12 marks)

Part 4: SQL (120 marks)

The three questions in this part of the TMA are based around Block 2 Parts 6–9. They focus on the Structured Query Language (SQL) – the standard language for defining, controlling and accessing relational databases.

Question 5 is worth 20 marks, Question 6 is worth 65 marks and Question 7 is worth 35 marks.

Question 5 can be answered after you have completed your study of Block 2 Part 6 and any related activities within that part. It is designed to allow you to demonstrate your understanding of the use of normalisation when checking a data model, and the process of translating between logical and physical schema representations.

Question 6 can be answered after you have completed your study of Block 2 Parts 7 and 8, and any related activities within that part. It is designed to allow you to demonstrate that you can write and interpret SQL queries to manipulate data in a given data model to meet stated requirements, and that you can relate an SQL query to the logical processing model.

Question 7 can be answered after you have completed your study of Block 2 Part 9 and any related activities within that part. It is designed to allow you to demonstrate your understanding of the access and security features of SQL and issues around shared data.

Question 5 (20 marks)

a. Before developing a database solution, the company had been using a spreadsheet to record basic information about events. A fragment of this spreadsheet is used in this question. The fragment shows the event id, the date of the event and the pilot, observer and drone used.

The company used the following relation as the template for the spreadsheet columns and rows shown below:

Event (EventID, Date, PilotEmployeeID, PilotName, ObserverEmployeeID, ObserverName, Manufacturer, SerialNumber)

The following are known to always hold between the attributes of the above relation:

A value of EventID uniquely determines a single Date value.

A value of EventID uniquely determines a single PilotEmployeeID value.

A value of EventID uniquely determines a single PilotName value.

A value of EventID uniquely determines a single ObserverEmployeeID value.

A value of EventID uniquely determines a single ObserverName value.

A value of EventID uniquely determines a single Manufacturer value.

A value of EventID uniquely determines a single SerialNumber value.

A value of PilotEmployeeID uniquely determines a single PilotName value.

A value of ObserverEmployeeID uniquely determines a single ObserverName value.

A table of sample data representing the relation is shown in Figure 5.

EventID	Date	PilotEmployeeID	PilotName	ObserverEmployeeID	ObserverName	Manufacturer	Serialnumber
E00282	10/09/19	002341	Joshua Smith	002346	Hannibal Heyes	RotorDyne	RD142562324
E00275	1/09/19	002330	Thaddeus Jones	002346	Hannibal Heyes	RotorDyne	RD142562324
E00271	1/08/19	002330	Thaddeus Jones	002341	Joshua Smith	FlyteSpeed	1242322
E00283	10/09/19	002330	Thaddeus Jones	002347	Jedediah Curry	RotorDyne	RD142566243
E00284	11/09/19	002341	Joshua Smith	002346	Hannibal Heyes	RotorDyne	RD142566243
E00285	11/09/19	002330	Thaddeus Jones	002347	Jedediah Curry	FlyteSpeed	1242322
E00290	15/09/19	002341	Joshua Smith	002346	Hannibal Heyes	FlyteSpeed	1442522
E00291	17/09/19	002330	Thaddeus Jones	002341	Joshua Smith	FlyteSpeed	1442522

Figure 5 The relation event represented by a table of sample data.

i. Using the data in Figure 5 describe, with an example, how an amendment anomaly might arise.

(4 marks)

ii. Briefly explain why Event is in first normal form.

(3 marks)

iii. Briefly explain why Event is in second normal form.

(2 marks)

iv. Event is not in third normal form. Identify the problems, and then split the relevant attributes into new tables so that the resulting tables are in third normal form.

(5 marks)

(14 marks)

b. Convert the relational representation shown in Figure 6 to appropriate SQL Data Definition Language statements to create tables representing the same data requirements.

You can assume that the referenced domains have already been defined in SQL as: CIDdomain, NameString, TelephoneNumber, EIDdomain, and EmployeeIDs, and the SQL DATE type is available. (The relational representation shown in Figure 6 uses a simplified version of the Event entity type and only the relationship between Event and Client is shown.)

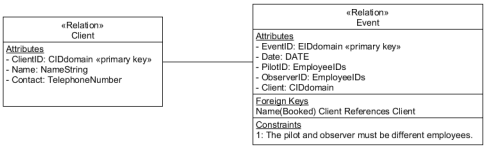

Figure 6 A relational model of the relationship Booked between the relations Client and Event.

(6 marks)

Question 6 (65 marks)

This question requires you to use the tables of the drone database described in the introduction to the TMA. You will need to install the tables and data by evaluating the drone.txt file in psql which will create tables according to the schema in figure 7 and insert data into these tables.

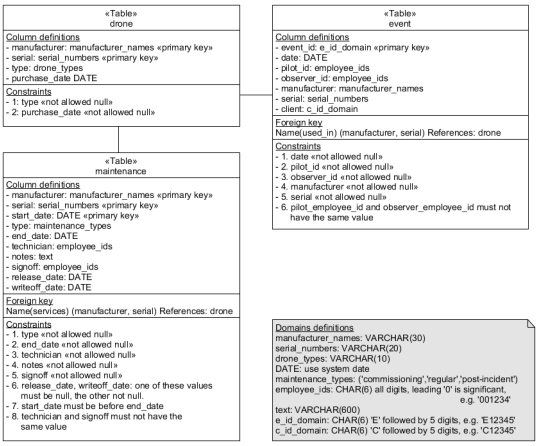

Figure 7 A schema for the tables defined in the drone.txt file.

In your solution document please include your SQL statements and a copy of the output produced by the evaluation of each statement.

a. Write SQL statements that answer each of the requests for data in the drone.txt tables:

   i. Produce a list of employee identifiers used in the company. The list should include the employees in the roles of pilot, observer, technician and signoff (technician); it should not include duplicate identifier values. (HINT a list without duplicates can be considered as a set of values).

   ii. For each event that is a training or test event, give the date, pilot_id, observer_id and manufacturer. (Note that training and test events have no client identified for that event.)

iii. How many maintenance records are there where the type was 'post-incident'?

iv. How many different manufacturers have produced at least one drone bought by the company?

v. List the date of the maintenance start, the manufacturer and serial number of drones at regular maintenance sessions where that maintenance session's notes include a mention of 'rotor'.

vi. For each event, give the event_id, manufacturer, serial number and type of the drone used at that event.

vii. List the manufacturer, serial number, start date and maintenance type of each maintenance session where the technician was either employee 025524 or 002347; the resulting list should be in date order.

viii. For each pilot in the event table, give the pilot identifier and a count of the number of events where they have piloted.

ix. List the drones owned by the company and how often they've been used in events. The list should show, for each drone, the manufacturer, serial number and the count of events that drone has been used at. If the drone hasn't been used at any events then it should still be listed, with a count of 0 (zero).

x. Which observers have been observers at fewer than 3 events?

xi. For each commissioning maintenance session, list the start date of the maintenance session and the manufacturer, serial and type of the drone being commissioned. The result should be presented in date order with the most recent session listed first.

(42 marks)

b. What request does the following SQL answer? (Note that a request should be an English language question or description like the ones in part a), *not* an account of how the data is processed to produce the result.)

```
SELECT drone.manufacturer,drone.serial,purchase_date,
writeoff_date
FROM drone JOIN maintenance
ON (drone.manufacturer, drone.serial) =
    (maintenance.manufacturer, maintenance.serial)
WHERE (writeoff_date - purchase_date) < 14;
```

(3 marks)

c. Describe, using the logical processing model, the evaluation of the following query:

```
SELECT manufacturer, count(start_date)
FROM maintenance
WHERE (end_date - start_date) > 3
GROUP BY manufacturer
HAVING count(start_date) > 1;
```

(8 marks)

d. Consider the information:

The FlyteSpeed drone, serial number 1442522 underwent maintenance following a collision with a Pelican. Maintenance is performed by employee 025525 between 1/03/2020 and 4/03/2020 and then signed off by employee 003451 on 5/03/2020. The maintenance identified structural failure of the

battery housing and it was not repairable. It was recorded as written off on 5/03/2020.

Add this information to the maintenance table using an `SQL INSERT` statement, then write a query to show that your insertion was successful.

(3 marks)

e. Briefly describe the similarities and difference between the NATURAL JOIN and INNER JOIN?

(4 marks)

f. The company wants to identify drones that have been purchased but not commissioned.

Create an SQL VIEW with the name `uncommissioned` that selects rows from the `drone` table where there is not a corresponding maintenance session. The view should show the manufacturer name, type, serial number and the date of purchase, for those rows selected. The columns in the view should have the names manufacture, type, serial, and date respectively.

Then, write a query showing the data the view contains.

(5 marks)

Question 7 (35 marks)

a. *You should write no more than **300 words** for this part of the question.*

The company is often told by clients that the work they are undertaking is commercially sensitive and the details of their bookings and events should not be disclosed to outside parties. The company require their database to be developed and managed in such a way that they have a high degree of confidence in the security of the client specific data (available to a limited few within the company) while also ensuring that the other data within the database is readily accessible by a wider range of company employees, for example to enter data into non-sensitive tables and extract data that does not contain sensitive information.

Considering the security and data management facilities for relational databases considered in the course, describe how the company could ensure the client information can be protected.

Your description should consider a client table (which would contain details of their clients) and the presence of the client identifier in the event table used in question 6. You should draw on the principle of least privilege and the rules for controlled sharing when answering this question.

(15 marks)

b. *You should write no more than **100 words** for this part of the question.*

Why is the data generated by the drones at these events not going to be a major part of the database security concerns discussed above; and what can the DBMS protect in relation to the client data files? (You may want to refer back to the Scenario Overview when answering this question.)

(5 marks)

c. *You should write no more than **300 words** for this part of the question.*

Alongside the company's historic record keeping system (described in the scenario) they are also developing a booking system for events.

A key element of this booking system is an application that will capture and record the details for a planned event. The application will capture the client details, details of the location the client wishes to use, the date for the booking and then identify, using knowledge of other bookings and availability records, which drones, pilots and observers can be allocated to the booking and are suitable for the task the client wishes to have completed. Once the application has assembled the required information for the booking a record for the booking is stored in a database, and updates the associated core tables: the client table if it is for a new client, the location table if it is a new location, the drone, pilot and observer availability, etc.

The company hopes to have several sales advisors who will be able to discuss requirements with potential and existing clients and make bookings directly using the above application – sales advisors using a paper system say it can take 10-20 minutes working with clients to complete booking information.

Briefly explain why the booking application described above would require a transaction management system capable of realising the ACID properties for transactions. You should include examples of the problems that ACID transaction management would prevent in the above example.

(15 marks)

WEB 05819 2

2.1