University of Brighton

SCHOOL OF COMPUTING, ENGINEERING AND MATHEMATICS

SEMESTER TWO 48 HOUR COURSEWORK 2019/20

CI330

Data Management

Time allowed: 48 hours

Answers: Complete on separate Answer Book provided, adding scans photographs if required

Answer All FOUR (4) questions on the paper in the answer document provided

Items permitted: There is no restriction on the paper-based items that the student may have available for the examination

Items supplied: Word Answer Sheet

All work submitted must be your own and all sources which do not fall into that category must be correctly attributed. The markers may submit the whole set of submissions to the JISC Plagiarism Detection Service.

Marks for whole and part questions are indicated in brackets ( )

 **The Welcome Hotel Group Case Study** 

The Welcome Hotel group operates 400 hotels across the world, in 32 different countries. It has a turnover of $100 million per year worldwide. As well as the hotels, there are area Regional Head Offices in:

• London, United Kingdom	Europe Headquarters
• Tokyo, Japan	Asia Headquarters
• Sydney, Australia	Oceania Headquarters
• Florida, United States of America	Americas Headquarters

You work in the Information Systems department in the London Head Office.

There is an enterprise-wide information system, covering the most areas of the business, including:

- Room and facilities booking
- Purchasing
- Pay and personnel (staff) records

The hotel information is stored on a distributed relational database system. Although there are many new applications that access the database, the core database design has remained the same for over ten years.

There are 3 hotels in the South-East of England with the following *hotelIDs.* These are:

- Brighton Welcome Hotel ("BN")
- Guildford Welcome Hotel ("GF")
- Winchester Welcome Hotel ("WC")

In the current information system, the *room bookings* are held locally in the hotel concerned. For example, the Welcome Hotel (Brighton), holds its own room and facilities bookings on their own local server. This table is called BOOKING(BN). Booking tables in other hotels follow this naming pattern.

The London Regional Head Office holds staff information for all hotels in Europe. Staff information about disability is considered *sensitive personal data.* The two parts of the STAFF table are called STAFF_MAIN and STAFF_DIS.

Consider the following (partial) data schema from the hotel system:

- HOTEL (hotelID, hotelName, hotelAddress, hotelTel, hotelCountry)
- ROOM (roomID, *hotelID, roomType, roomRate£)
- GUEST (guestID, guestFirstName, guestSurname, guestContactNo, guestAddress)
- BOOKING (*roomID, *hotelID, *guestID, bookingStartDate, noOfNights, specialRequirements)
- STAFF (staffID, *hotelID, staffGrade, salary, staffFName, staffSName, NINumber, dateStarted, disabilityInformation)

Question 1: Distributed Databases

(a) Give TWO (2) reasons why the Welcome Hotels hold *booking information* on their own local servers, rather than in a central server for all hotels. Illustrate your answer with specific reference to the Welcome Hotel Case Study.

(6 marks)

(b) Write the SQL to find the staff who work in the Brighton Welcome Hotel from the staff table held in the Regional Head Office in London. This created table will be called STAFF(Brighton).

(4 marks)

(c) In the above case study, the BOOKING table is *distributed horizontally,* and the STAFF table is *distributed vertically*. Explain what is meant by *horizontal* and *vertical distribution*, using these as examples. Write the relational algebra to re-unite (join together) these fragments to show the whole STAFF table (include disability data) *and* BOOKINGs for the THREE (3) South-East England hotels.

(10 marks)

(d) A distributed database management system (DDBMS) needs additional functionality compared to a one-site database management system (DBMS). Identify TWO (2) additional functions required for a DDBMS. For *each* named function, explain what it means, illustrating with an example from the Welcome Hotel Case Study.

(8 marks)

(28 marks total)

Question 2: Security and Legal Issues

You have a junior colleague from the United States of America who does not know the law covering personal data in the UK. Provide short notes for her, answering the following questions:

(a) Which law covers personal data processing in the UK and European Union (EU)?

(2 marks)

(b) What is *personal data*? Include an example from the case study.

(5 marks)

(c) What is *special category data*? Include an example from the case study.

(5 marks)

(d) A staff member has accidently published names and addresses of all guests on the Internet. This is known as a *data breach.* What should the Welcome Hotel Group do to comply with UK law?

(6 marks)

(18 marks total)

Question 3: Non-Relational Databases, Database Selection

(a) The Welcome Hotel Group currently uses a relational database system to store BOOKING data. Identify ONE (1) non-relational data storage model. Use the BOOKING data to illustrate how that storage model works.

(11 marks)

(b) Suggest TWO (2) possible advantages of storing the data in the non-relational format that you have identified in part (a) of this question, with specific reference to the Welcome Hotels Case Study.

(6 marks)

(c) Give TWO (2) reasons why the Welcome Hotel Group may decide to continue with the current Relational database.

(4 marks)

(21 marks total)

Question 4 – Query Optimisation and Transaction Management

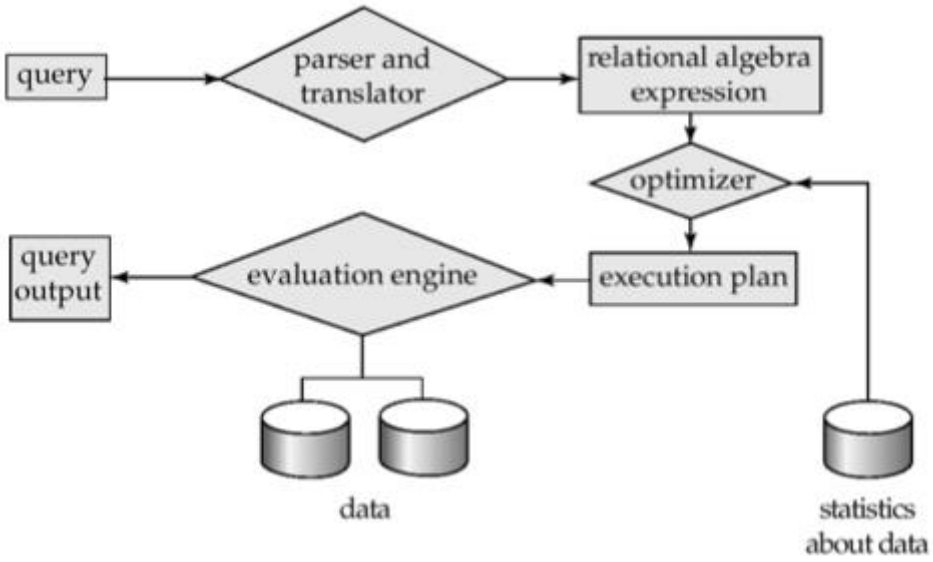

Figure 1: Query Processing - Relational Database

(a) Use the following SQL as an example to explain the main stages of query processing as shown in *figure 1.*

*Select * from GUEST inner join BOOKING*
on GUEST.guestID = BOOKING.guestID
where guestSurname = "Samuels"
order by bookingStartDate

Please state any assumptions that you make about the data / database structure in your answer.

(10 marks)

[Question 4 continues on next page]

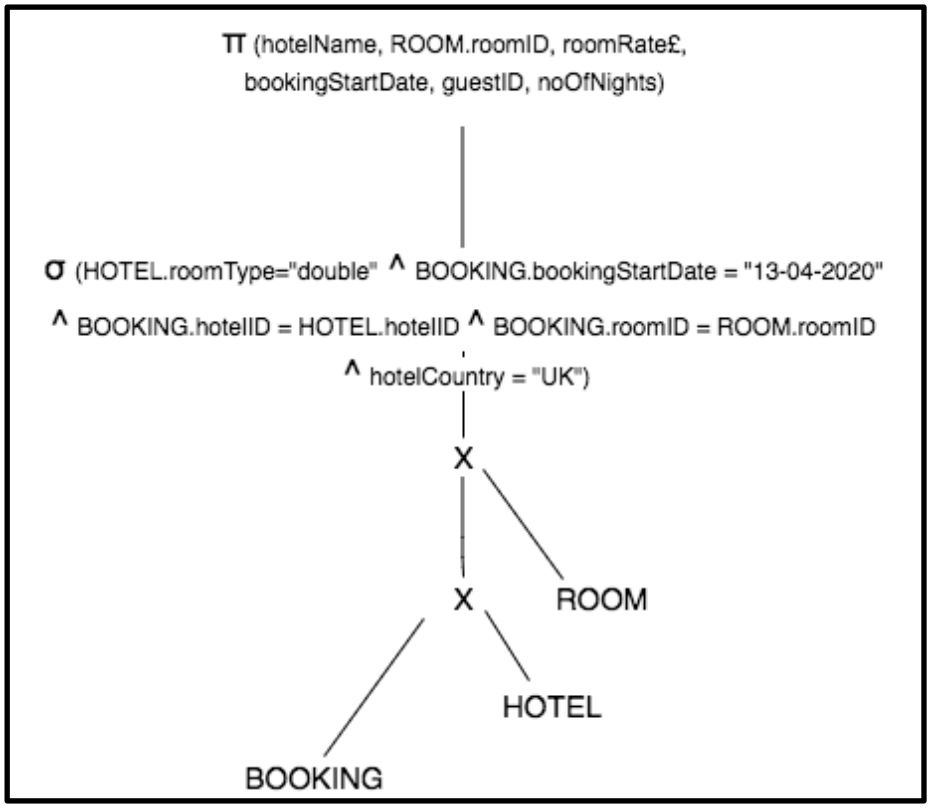

Figure 2: Relational Algebra Tree (RAT)

(b) Examine the relational algebra tree (RAT) in *figure 2*. Optimize the RAT to produce a more efficient execution, using heuristic optimization.

Consider the following statistics about the Welcome Hotel database system to inform your answer:

- There are 400 hotels, of which 15% are in the UK
- Hotels have an average of 100 rooms in them
- 75% of rooms are booked per night
- 50% of rooms are of roomType = “double”

(13 marks)

[Question 4 continues on next page]

(c) The following table contains two booking transactions, *booking1* and *booking2* which run concurrently, one trying to book two rooms, one trying to cancel a room for the Brighton Welcome Break Hotel for tonight. The value *Number of Remaining Rooms (rr)* is a running total – when it is more than zero, the hotel knows that it has rooms still available to be booked for tonight – this can be the basis for late booking deals.

Time	booking1 *(making booking)*	Booking2 *(cancelling booking)*	Number of remaining rooms for tonight (rr)
T1	Begin Transaction		2
T2	Read (rr)	Begin Transaction	2
T3	rr = rr - 2	Read (rr)	2
T4	Write (rr)	rr = r + 1	0
T5	Commit	Write (rr)	3
T6		Commit	3

(i) Name the problem shown here, and state how many *remaining rooms* for tonight should the hotel now be showing?

(2 marks)

(ii) Redraw the table above, using two-phase locking, to solve this problem.

(8 marks)

(Question Total: 33 marks)

[End of paper]