

IFN554 Databases Assignment (Part 2)

Weighting: 50%

Due date for final submission: Friday April 3 at 11:59 pm (end of week 6)

Instructions

- This assignment is an individual task. You must NOT work in groups/teams.
- For this assignment, you must use **SQLite** DBMS. For Task2-5, you must use **Hotel** database.
- Assignment submissions are to be made via IFN554 Blackboard.
- Assignment submissions MUST be in a **single PDF file**.

Late submission

According to QUT policies, submissions past the deadline will NOT be marked, and will thus attract a mark of 0. Therefore, you are strongly advised to submit your report, even if it is only partially complete, by the due date.

Extensions

According to QUT policies, an extension to the due date of the assignment may only be granted on medical/compassionate grounds. Please see the *Late assignments and extensions information on the unit's Blackboard Assessment* page.

Reviews

We cannot remark assessments. All assessments will be returned with feedback explaining the reasons for the marks allocated. If you require additional feedback or clarification, discuss it with your tutor. If you believe that there is a component that has not been marked in accordance with the criteria sheet, you must identify it in writing to your tutor.

Academic honesty

Any action or practice on your part which would defeat the purposes of assessment is regarded as academic dishonesty. The penalties for academic dishonesty are provided in the Student Rules. For more information consult the QUT Library resources for avoiding plagiarism.

Getting feedback

The teaching team will be available to answer specific questions about the assignment, but **not to pre-mark** assignments. Lecturers and tutors will **NOT** read report drafts and review detailed models prior to the submission of the assignment.

Detailed feedback will be provided in written form when the marker returns the marked assignment back to the students. Students are also welcome to discuss any queries about the feedback they will receive.

General inquiries

For specific content-related inquiries about your assessments, please contact your tutor.

Task 1 [6 marks]

Write an SQL script that builds a database to match the relational model provided to you. These SQL statements in the script must be provided in the correct order.

You are required to create a database for the fictitious bookstore Oktomook for Task 1. The database is based on the following model.

OKTOMOOK RELATIONAL MODEL

Branch (branchNumber, branchName, streetNo, streetName, branchCity, branchState, numberEmployees)

Publisher (publisherCode, publisherName, publisherCity, publisherState)

Author (authorID, firstName, lastName)

Book (ISBN, title, publisherCode, genre, retailPrice, paperback)

Wrote (ISBN, authorID)

Inventory (ISBN, branchNumber, quantityInStock)

FOREIGN KEYS

- Book(publisherCode) is dependent on Publisher (publisherCode)
- Wrote (ISBN) is dependent on Book (ISBN)
- Wrote (authorID) is dependent on Author (authorID)
- Inventory (ISBN) is dependent on Book (ISBN)
- Inventory (branchNumber) is dependent on Branch (branchNumber)

OTHER CONSTRAINTS and REMARK

- branchNumber comprises three digits.
- ISBN comprises 10 digits.
- publisherCode comprises one letter and two digits.
- authorID comprises four digits.
- INTEGER type must be used for retailPrice, ~~numberEmployees~~ and quantityInStock.
- TEXT type must be used for other attributes (except retailPrice, numberEmployees and quantityInStock).
- Book title always contains a value.
(i.e., you cannot insert a new record, or update a record without adding a value to this field)

Write an SQL script for creating new tables including all attributes.

Marks will be awarded for the following:

- Successfully creating new tables (2 marks)
- Including all attributes and correct data types (2 marks)
- Correctly creating Primary Keys (1 mark)
- Correctly creating Foreign Keys (1 mark)
- Correctly creating other constraints (1 mark)

Create table Branch

```
...> (  
...> branchNumber Text Primary Key Not Null Check(Length(branchNumber)=3),  
...> branchName Text,  
...> streetNo Text,  
...> streetName Text,  
...> branchCity Text,  
...> branchState Text,  
...> numberEmployees Integer  
...> );
```

Create table Publisher

```
(  
  publisherCode Text Primary Key Not Null Check(publisherCode GLOB '[a-zA-Z][0-9][0-9]'),  
  publisherName Text,  
  publisherCity Text,  
  publisherState Text  
);
```

Create table Author

```
...> (  
...> authorID Text Primary Key Not Null Check(Length(authorID)=4),  
...> firstName Text,  
...> lastName Text  
...> );
```

Create table Book

```
...> (  
...> authorID Text Primary Key Not Null Check(Length(authorID)=4),  
...> title Text,  
...> publisherCode Text Check(publisherCode GLOB '[a-zA-Z][0-9][0-9]'),  
...> genre Text,  
...> retailPrice Integer,  
...> paperback Text,  
...> foreign key (publisherCode) references Publisher(publisherCode)  
...> );
```

Create table Wrote

```
...> (  
...> ISBN Text Check(Length(ISBN)=10),  
...> authorID Text Check(Length(authorID)=4),  
...> primary key (ISBN, authorID),  
...> foreign key (ISBN) references Book (ISBN),  
...> foreign key (authorID) references Author(authorID)  
...> );
```

Create table Inventory

```
...> (  
...> ISBN Text Check(Length(ISBN)=10),  
...> branchNumber Text Check(Length(branchNumber)=3),  
...> quantityInStock Integer,  
...> primary key (ISBN, branchNumber),  
...> foreign key (ISBN) references Book (ISBN),  
...> foreign key (branchNumber) references Branch(branchNumber)  
...> );
```

Task 2 [15 marks]

We have provided you the Hotel database to be used with SQLite DBMS. You should use this database in SQLite to extract the necessary information as per the following query requirements.

The script is based on the following relational schema:

- Hotel (hotelNo, hotelName, city)
- Room (roomNo, hotelNo, type, price)
- Booking (hotelNo, guestNo, dateFrom, dateTo, roomNo)
- Guest (guestNo, guestName, guestAddress)

Note the following details of the hotel database;

- Hotel contains hotel details and hotelNo is the primary key;
- Room contains room details for each hotel and (roomNo, hotelNo) forms the primary key;
- Booking contains details of bookings and (hotelNo, guestNo, dateFrom) forms the primary key;
- Guest contains guest details and guestNo is the primary key.

Write an SQL script for querying data.

- List hotelNo, type and price of each double or deluxe room with a price more than \$99 (2 marks).

Select hotelNo, type, price from Room where (type="Double" or type="Deluxe") and (price>99);

H2/Double/100

H2/Deluxe/144

H3/Deluxe/132

H5/Double/100

H5/Double/100

H6/Double/100

- List hotelNo who have more than 2 double rooms (2 marks).

Select hotelNo from Room where type="Double" Group By hotelNo Having Count(*)>2;

H5

- List number of different guests who visited Ridge Hotel (3 marks).

Select Distinct b.guestNo, a.hotelName from Hotel a join Booking b on a.hotelNo = b.hotelNo where a.hotelName = "Ridge Hotel";

G1/Ridge Hotel

G3/Ridge Hotel

- What is the total income from bookings for the Grosvenor Hotel? (4 marks).

Select Sum((julianday(b.dateTo) - julianday(b.dateFrom))) * (c.price) AS total_pay from Hotel a, Booking b, Room c on a.hotelNo = b.hotelNo where (c.hotelNo = b.hotelNo and c.roomNo = b.roomNo) and a.hotelName = "Grosvenor Hotel";

6560.0

- List all the guests who have stayed in a hotel (4 marks).

Select Distinct(b.guestName) from Booking a, Guest b where (b.guestNo = a.guestNo);

David

Robert

Esther

Hari

I believe the above is correct, however, it probably makes more sense to check whether or not the booking has incurred any cost, as these shouldn't be recorded as a stay. Considering this, the below code would be correct:

```
Select Distinct(d.guestName) from Hotel a, Booking b, Room c, Guest d on a.hotelNo = b.hotelNo  
where (c.hotelNo = b.hotelNo and c.roomNo = b.roomNo and d.guestNo = b.guestNo) and (b.dateTo  
> b.dateFrom);
```

**David
Esther
Robert
Hari**

Marks will be awarded for the following:

- Full marks will be awarded for each query, if the query produces the correct output.

Task 3 [3 marks]

Perform the following tasks.

- Write commands to insert rows in each of the Hotel database tables (1 marks).

Insert into Hotel

Values("H8", "Meriton Hotel", "Brisbane");

Insert into Room

Values("R1", "H8", "Single", 85);

Insert into Booking

Values("H8", "G1", "2010-02-15", "2010-02-19", "R1");

Insert into Guest

Values("G7", "Chris", "Brisbane");

- Write a command to delete the row you inserted in the table Guest (1 marks).

Delete from Guest

where guestNo = "G7";

- Write a command to update the price of all rooms by 10% (1 mark).

Update Room set price = price *1.10;

Marks will be awarded for the following:

- Full marks will be awarded for each query, if the query produces the correct output.
- As four Insert commands are required, 0.25 mark for each working command will be given. (0 mark if an Insert command is not able to insert a row).
- 1 mark if the command is able to delete the row otherwise 0 mark.
- 1 mark if the command is able to update the price by 10% otherwise 0 mark.

Task 4 [2 marks]

Perform the following tasks.

- Currently the database only contains a small number of records, however the data contained within it is expected to grow significantly in the future. Creating indexes on commonly searched columns is a way performance issues can be minimised. Write a command to create an index on guestName of the Guest table (1 mark).

Create index guest_index on Guest(guestName);

- Write a command to create a view to list the information (hotelName, roomType and the total number of rooms booked) of the hotels which are in Cairns. (1 mark).

Create view Question_4_Info as Select b.hotelName, room.type, Count(*) From Room, Hotel b Join Booking on (booking.roomNo = room.roomNo and booking.hotelNo = room.hotelNo and booking.hotelNo = b.hotelNo and b.city = "Cairns") group by room.hotelNo, room.type;

Marks will be awarded for the following:

- Full marks will be awarded for each query, if the query produces the correct output. Otherwise 0 mark.

Task 5 [4 marks]

Nikki and Phil work with the Hotel database as database administrator. Provide the commands required to grant or revoke access so the following security requirements are met:

Perform the following tasks.

- User Nikki must be able to add records to the Booking table (1 Mark)
- User Nikki must be able to remove records from the Booking table (1 Mark)
- User Phil is no longer allowed to add data to the Guest table (1 Mark)
- User Phil is no longer allowed to delete records from the Guest table (1 Mark)

Assume usernames of employees Nikki and Phil are nikki and phil respectively.

CREATE USER 'nikki' identified by 'administrator'
CREATE USER 'phil' identified by 'administrator'

(Above is only relevant if the user haven't already been created).

GRANT INSERT ON Booking to nikki;
GRANT DELETE ON Booking to nikki;
REVOKE INSERT ON Guest from phil;
REVOKE DELETE ON Guest from phil;

Note: You will not be able to try these commands in SQLite.

Marks will be awarded for the following:

- Full marks will be awarded for each query, if the query is correct. Otherwise 0 mark.

Task 6 [10 marks]

Write a 300 – 500 words report on the possible considerations for ethical use of data from the perspective of data management.

Marks will be awarded for the following:

- The key considerations for ethical use of data are clearly articulated and comprehensively discussed & Excellent formatting and impeccable grammar. (10-8 Marks)
- The key considerations for ethical use of data are clearly identified and discussed & Good formatting and very minor grammatical errors. (7-6 Marks)
- The key considerations for ethical use of data are adequately identified and discussed & Satisfactory formatting and a few grammatical errors. (5-4 Marks)
- Some of the key considerations for ethical use of data are identified and discussed & Poor formatting and many grammatical errors. (3-2 Marks)
- None of the key considerations for ethical use of data are identified or discussed & Non-existent formatting and rife with fundamental grammatical errors. (1-0 Marks)

The purpose of this report is to improve the literacy around the ethical use of data within the realm of data management. This report covers four key factors; user consent, secondary use, privacy violation and data quality. To help with the understanding of these concepts, we will be using a very simple use case of buying a product online. In each section of this report, the key consideration will be explained and a link to the previously explained scenario will be given.

The first key consideration is user consent. User consent, as it suggests, refers to a user consenting to their data being collected, processed or stored. The “General Data Protection Regulation” (GDPR) is one regulation example which enforces organisations to obtain explicit permission, rather than hiding the consent in hard to read documents (Auvray & Podnar, 2018). An example in our online purchase scenario would be a website providing a button to opt in to receiving their marketing material. If they were to act unethically (and potentially illegally), they would just collect your email and send their marketing material to you.

The second consideration is secondary use. This key consideration refers to organisations only using the data for the purpose it was collected for. An example using our scenario would be a website asking for your address to ship your purchase to, but then using your address (and other addresses) to send targeted marketing material to. This ethical area of data management is particularly prevalent in the health industry.

The third consideration is privacy violation. Data privacy somewhat overlaps with the first consideration of user consent; however, it focuses more specifically on policies and documents that govern how data is collected, shared and used. An example from the GDPR is the right to not be subjected to decisions formed from automated data processing (Peters, 2020). An example from our scenario would be the website using an algorithm to determine what the maximum price that someone from our demographic would pay and then adjusting the price to this value.

The last key consideration is data quality. This consideration focuses on the balance between having high quality data in order to provide high quality services vs having data which passes the data privacy regulations highlighted previously. For example, an online shop may believe that by creating an algorithm that determines products that I'm mostly likely to buy and then market to me, might be the best outcome for me, however, this may breach data privacy regulations.

In conclusion, this report has clearly highlighted and explained the key considerations for ethical use of data. From this report, readers should better understand the importance of ethical considerations and should start to shift towards having these ethical decisions at the forefront during their decision making.

References

- Auvray, E & Podnar, K (2018) *An Introduction to the GDPR*. <https://www.cmswire.com/customer-experience/an-introduction-to-the-gdpr/>
- Peters, J (2020) *Data Privacy Guide: Definitions, Explanations and Legislation*. <https://www.varonis.com/blog/data-privacy/>

Task 7 [10 marks]

Using the following table structure, identify all functional dependences and then decompose this table into a set of 3NF relations.

Assumptions

In the following table, these assumptions can be made:

1. there are no multivalued dependencies
2. any invoice numbers may reference more than one product
3. any given product is supplied by a single vendor, but a vendor can supply many products

INV_NUM	PROD_NUM	CUS_NUM	PROD_DESCRIPTION	VEND_CODE	NUMBER_SOLD	PROD_PRICE
211347	AA-E3422QW	C001	Rotary Sander	211	1	\$49.95
211347	QD-300932X	C001	0.25-in. drill bit	211	8	\$3.45
211347	RU-995748G	C001	Band Saw	309	1	\$39.99
211348	AA-E3422QW	C004	Rotary Sander	211	2	\$49.95
211349	GH-778345P	C004	Power Drill	157	1	\$87.75

Whole Key

Is CUS_NUM a fact about the whole key? No, it's about INV_NUM or Prod Num.

Is Prod description a fact about the whole key? No, it's about Prod_Num

Is Vend_Code a fact about the whole key? Yes.

Is Numbers_Sold a fact about the whole key? Yes

Is Prod_Price a fact about the whole key? No, it's about Prod Num.

2NF

INV_NUM	PROD_NUM	VEND_CODE	NUMBER_SOLD
211347	AA-E3422QW	211	1
211347	QD-300932X	211	8
211347	RU-995748G	309	1
211348	AA-E3422QW	211	2
211349	GH-778345P	157	1

Is Vend code a fact about number sold? No

Is Number Sold a fact about Vend Code? No

Is prod Num a fact about vend code? No

Is vend code a fact about prod num? Yes

3NF

INV_NUM	PROD_NUM	NUMBER_SOLD
211347	AA-E3422QW	1
211347	QD-300932X	8
211347	RU-995748G	1
211348	AA-E3422QW	2
211349	GH-778345P	1

3NF

PROD_NUM	VEND_CODE
AA-E3422QW	211
QD-300932X	211
RU-995748G	309
GH-778345P	157

3NF

PROD_NUM	CUS_NUM	PROD_DESCRIPTION	PROD_PRICE
AA-E3422QW	C001	Rotary Sander	\$49.95
QD-300932X	C001	0.25-in. drill bit	\$3.45
RU-995748G	C001	Band Saw	\$39.99
GH-778345P	C004	Power Drill	\$87.75

Is CUS_NUM a fact about Prod Description? No

Is Product description a fact about CUS_NUM? No

Is CUS_Num a fact about Prod_Price? No

Is Prod_Price a fact about Prod Description? No

Appendix A – Declaration template

By submitting this assignment, I am aware of the University rule that a student must not act in a manner which constitutes academic dishonesty as stated and explained in the QUT Manual of Policies and Procedures. I confirm that this work represents my individual effort. I declare that it does not contain plagiarized material.

Full Name	Student No.	Signature
Christopher Dugdale	N10572988	Christopher Dugdale