COMP2400/6240 – Relational Databases
Semester 2, 2024

Assignment 2 (Database Theory)

Due date: 17:00 (Canberra time), 11 October 2024

Instructions

- **This assignment must be done individually.**

  Do not share your solutions, or partial solutions, with anyone. Do no post any idea/partial solution/result related to the assignment on the course discussion forum, or anywhere else where it can be read by other students.

  You may of course ask clarification questions, but make sure you phrase your questions so that they do not give away what you think the answer or solution is. If you are not sure if a question is ok to post to the forum then ask your tutor during a lab or come to a drop-in session instead.

  You must not use generative tools (ChatGPT or similar) to create your answers to the assignment. Your answers should be *your* answers. Reference all resources that you have used in completing this assignment.

- **You must submit your solutions through Wattle before the assignment deadline.** There will be an Assignment activity on the course Wattle page where you should upload your solution file. You can submit more than once, but we can only see, and will only mark, the last file that you submit.

  Late submissions will *not* be accepted. Submissions made outside the Assignment activity on Wattle (for example, files sent by email) will *not* be accepted under any circumstance. You will be marked on what you have submitted at the time of the deadline.

- This assignment will count for 17% of the final course mark. Marks are reserved for detailing the process of deriving your solution, not just for the result. You should justify your answers and include all essential ideas and steps to derive your solutions.

- You should try your best to type the solutions. The scanned images of handwritten texts and equations can be unreadable for marking. As for the EER diagram, you are recommended to export a JPEG file from TerraER and include it in the PDF file.

- **Plagiarism, collusion, and the use of disallowed tools will attract academic penalties in accordance with the ANU guidelines. Every student in this course is expected to be able to explain and defend any submitted assessment item. The course conveners can conduct or initiate an additional interview about any submitted assessment item for any student. If there is a significant discrepancy between the two forms of assessment, this will be seen as a case of potential academic misconduct.**

- If you find yourself in an unforeseeable situation beyond your control that you believe significantly affects your ability to complete the assessment on time, you can submit an ECA through the system (`https://www.anu.edu.au/students/program-administration/assessments-exams/extenuating-circumstances-application`).

Question 1 **6 Marks**

FlyHigh is an airline company that specializes in providing on-demand charter flight services. The company operates a fleet of ten aircraft, catering to a diverse range of customers who require flexible and personalized flight options. To effectively manage its operations, FlyHigh maintains records for each customer, including their unique customer number, full name, contact phone number, and residential or business address.

To ensure the reliability and safety of its fleet, FlyHigh has established five dedicated service teams. Each service team is composed of three service engineers and one administrative staff member. The service engineers are responsible for performing all necessary mechanical services on the aircraft, while the administrative staff handles the coordination and documentation of these services. FlyHigh's service teams can perform mechanical service on any of the aircraft in the fleet. To maintain optimal performance and compliance with safety standards, each aircraft undergoes a thorough mechanical service every six months.

FlyHigh classifies its employees into four categories: pilots, service engineers, flight attendants, and administrative staff. For each employee, the company maintains records that include their employee number, first name, last name, and date of birth. In addition to this general information, specific data is recorded for pilots, including their license number and category ratings.

The company also tracks detailed information about its aircraft. This includes the aircraft's registration number, model code, number of seats, mileage, engine strength, and manufacturer. For each aircraft's mechanical service, FlyHigh records the type of service performed, the date of the service, a description of the service, and any relevant comments.

When it comes to flight operations, an aircraft can be chartered by only one customer for a specific trip. Each flight requires a pilot and a co-pilot to operate the aircraft. All pilots must hold either a Commercial Pilot License (CPL) or an Air Transport Pilot License (ATPL). Furthermore, a pilot is only allowed to fly aircraft that fall within their specific category ratings, ensuring they are qualified and authorized to operate those types of aircraft. The company maintains records for each charter trip. The data held on each charter trip is the date of trip, start time, customer number, and aircraft registration number.

Your tasks are as follows:

1.1 Design an Enhanced Entity-Relationship (EER) diagram for the database application described above, using the notations from the lecture slides. Your diagram should include entities, relationships, attributes, and constraints where applicable. You should also document any assumptions you make while creating the EER diagram.

(4.5 Mark)

1.2 Identify any requirements in the database application described above that cannot be represented in an EER diagram.

(1.5 Mark)

Question 2 **5 Marks**

A job recruitment agency requested an information system to manage candidates, employers, and job applications. In response, the IT department developed a database schema that includes the following relation schema: JOB_APPLICATION.

JOB_APPLICATION

- Candidate_Name
- Candidate_DoB
- Candidate_Email
- Candidate_Education
- Job_ID
- Position
- Type
- Salary
- Location
- Closing_Date
- Employer_ID
- Employer_Name
- Employer_Address
- Application_Date
- Status

The IT department identified the data requirements from the job recruitment agency using the following functional dependencies:

(FD1) $\{Candidate\_Name, Candidate\_DoB\} \rightarrow \{Candidate\_Email, Candidate\_Education\}$;

(FD2) $\{Job\_ID\} \rightarrow \{Position, Type, Closing\_Date,\ Employer\_ID\}$;

(FD3) $\{Position, Type, Location, Employer\_ID, Employer\_Address\} \rightarrow \{Salary\}$;

(FD4) $\{Employer\_ID\} \rightarrow \{Employer\_Name, Employer\_Addresss\}$;

(FD5) $\{Employer\_Address\} \rightarrow \{Location\}$;

(FD6) $\{Candidate\_Name, Candidate\_DoB, Job\_ID\} \rightarrow \{Application\_Date, Status\}$.

Your tasks are as follows:

2.1 Based on the functional dependencies (FD1-FD6) provided, determine whether the relation schema JOB_APPLICATION is in 3NF. If it is not, apply the 3NF decomposition algorithm from the lecture notes to identify a 3NF decomposition for JOB_APPLICATION. To achieve full marks, ensure you include all key steps used in the 3NF decomposition process. (2 Mark)

2.2 Based on the functional dependencies (FD1-FD6) provided, determine whether the relation schema JOB_APPLICATION is in BCNF. If it is not: (3 Mark)

(a) Identify how many different BCNF decompositions exist for JOB_APPLICATION by applying the BCNF decomposition algorithm from the lecture notes. You must provide sufficient information to justify your answer. Simply giving a number without justification will not receive any marks.

(b) Present all BCNF decompositions that are both lossless and dependency-preserving. To achieve full marks, ensure you include all key steps used in the BCNF decomposition process.

Note that: (1) a 3NF/BCNF decomposition here means *the set of relation schemas and functional dependencies* you obtain after applying the 3NF/BCNF decomposition algorithms on JOB_APPLICATION (2) A simple answer "yes/no" without proper justification would not receive any mark.

Question 3 **6 Marks**

The following table contains the relational algebra operators covered in our course. You may only use these operators to answer the following questions.

$\sigma_\varphi R$	Selection by condition φ
$\pi_{A_1,\dots,A_n} R$	Projection onto the set of attributes $\{A_1 \dots, A_n\}$
$\rho_{R'(A_1,\dots,A_n)} R$	Renaming the relation name to $R^{'}$ and attribute names to $A_1, \dots, A_n$
$\rho_{R'} R$	Renaming the relation name to $R^{'}$
$\rho_{(A_1,\dots,A_n)} R$	Renaming the attribute names to $A_1, \dots, A_n$
$R_1 \cup R_2$	Union of two relations R_1 and R_2
$R_1 \cap R_2$	Intersection of two relations R_1 and R_2
$R_1 - R_2$	Difference of two relations R_1 and R_2
$R_1 \times R_2$	Cartesian product of two relations R_1 and R_2
$R_1 \bowtie_\varphi R_2$	Join of two relations R_1 and R_2 with the join condition φ
$R_1 \bowtie R_2$	Natural join of two relations R_1 and R_2
$\varphi_1 \wedge \varphi_2$	condition φ_1 AND condition φ_2
$\varphi_1 \vee \varphi_2$	condition φ_1 OR condition φ_2

Consider a simple hotel booking system with a relational database schema as follows:

- STAFF={eno, ename, ird} with the primary key {eno},
- ROOM={rno, location, facilities} with the primary key {rno},
- GUEST={gname, contact, phone} with the primary key {gname, contact},
- OFFER={rno, date, price} with the primary key {rno, date} and the foreign key:
  [rno] ⊆ ROOM[rno]
- BOOKING={gname, contact, rno, date, eno} with the primary key {gname, contact, rno, date} and with the foreign keys:
  [gname, contact] ⊆ GUEST[gname, contact]
  [rno, date] ⊆ OFFER[rno, date]
  [eno] ⊆ STAFF[eno]

We have the following relations over the above relational database schema:

STAFF

eno	ename	ird
1	Daffy	875-649-322
2	Walder	858-531-989

ROOM

rno	location	facilities
11	Level 0	Spa, Tennis court
13	Level 13	13 Bowling lanes
17	Level 2	Pool, Pool table

GUEST

gname	contact	phone
Pepe	Hotsforskunksville	0800 443344
Taz	Bunkerhill	0800 232323
Coyote	Ringroad	0800 838383

OFFER

rno	date	price
11	29/04/2016	749.00
13	29/04/2016	1313.13
17	29/04/2016	235.00
11	13/04/2016	749.00
13	13/04/2016	1313.13

BOOKING

gname	contact	rno	date	eno
Pepe	Hotsforskunksville	11	29/04/2016	1
Coyote	Ringroad	17	29/04/2016	1
Pepe	Hotsforskunksville	13	13/04/2016	1

Consider the following relational algebra query:

$$\pi_{gname,rno,location}(\sigma_{eno \neq eno' \vee rno \neq rno' \vee date \neq date'}(\text{ROOM} \bowtie \text{OFFER} \bowtie (\text{BOOKING} \bowtie \rho_{gname,contact,rno',date',eno'}(\text{BOOKING})) \bowtie \text{GUEST} \bowtie \text{STAFF}))$$

Your tasks are to

3.1 Write a SQL query that can be translated into the given relational algebra query.

(1 Mark)

3.2 Draw the query tree for the given relational algebra query.

(1 Mark)

3.3 Evaluate the relational algebra query over the given relations, and show each of your answers as a table that includes the attribute names and tuples. For example, such a table can be shown as follows:

gname	contact	rno	date
Pepe	Hotsforskunksville	11	29/04/2016
Coyote	Ringroad	17	29/04/2016
Pepe	Hotsforskunksville	13	13/04/2016

(a) Show the result of $\text{BOOKING} \bowtie \rho_{gname,contact,rno',date',eno'}(\text{BOOKING})$.

(c) Show the result of $\pi_{gname,rno,location}(\sigma_{eno \neq eno' \vee rno \neq rno' \vee date \neq date'}(\text{ROOM} \bowtie \text{OFFER} \bowtie (\text{BOOKING} \bowtie \rho_{gname,contact,rno',date',eno'}(\text{BOOKING})) \bowtie \text{GUEST} \bowtie \text{STAFF}))$.

(2 Mark)

3.4 Optimize the given relational algebra query. Your marks will depend on how well you present the key ideas of query optimization in your answer.

(2 Mark)