The Hong Kong Polytechnic University

SEMESTER I EXAMINATION 2022-2023

COMP 2411 – Database Systems

18th Oct. 2022 TIME ALLOWED: 2 HOURS

INSTRUCTIONS TO CANDIDATES

1. This is an **2-HOUR** exam.
2. The question paper contains **6** questions.
3. You need to answer all questions. The marks for each question are indicated at the beginning of each question.
4. This is **a CLOSE BOOK/NOTES** exam.

Student ID:

Name:

Question 1. [ER Diagram] (20 marks)

1. PolyU wants to create a very simple database for in which to record information about professors, students and classes. Please design an ER diagram for the database according to the following description. (10 marks)

   - For each professor we need to store the HK-id, name and office number.
   - For each student we need to store the student-id and name.
   - For each class the id (e.g., COMP 2411) and the name.
   - Each class is taught by exactly one professor.
   - Each student must take at least one class.
   - For each class that a student took we need to store the grade.

   Simplifying assumptions: there is only one lecture for each class and only one semester in the database.

   (a) Please give the three entity sets in this ER Diagram. (3 marks)

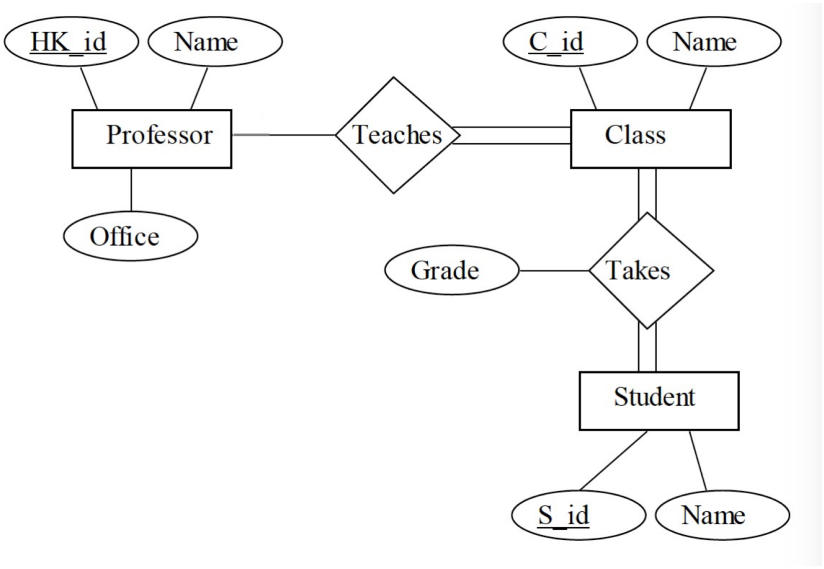
   Answer: professors, students and classes

   (b) Please give the two relationship sets in this ER Diagram. (2 marks)

   Answer: teach, take

   (c) Please draw an ER diagram for this database and please indicate keys and participation constraints in the diagram with their corresponding symbols. But you don't need to show the cardinality ratio. (5 marks)

   Answer:



2. A bus company wants to keep track of its bus routes and schedules. Design an ER diagram for the database according to the following description. (10 marks)

   - Each bus route has a route number, a departure station and a destination station.
   - For each bus route, there is a schedule, which records the departure times of buses.
   - For each departure time of each route, a driver and a bus can be assigned (however this is not necessary - information about the driver or the bus may sometimes be missing).
   - A driver has an employee Id, a name and a phone number.
   - A bus is identified by its license number. The database also records the seating capacity of each bus.

   (a) Please give the four entity sets in this ER Diagram. (4 marks)

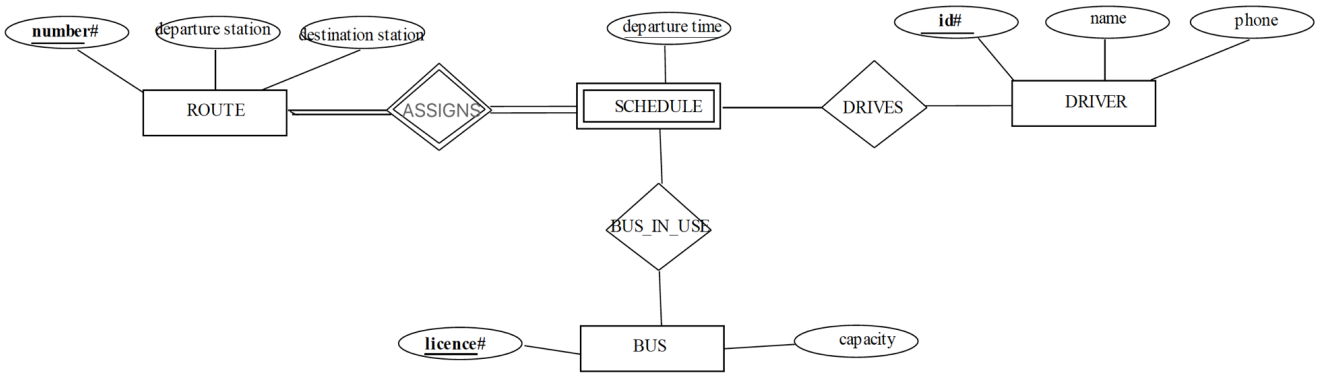
   Answer: route, schedule, driver, bus

(b) Please give the three relationship sets in this ER Diagram. (2 marks)

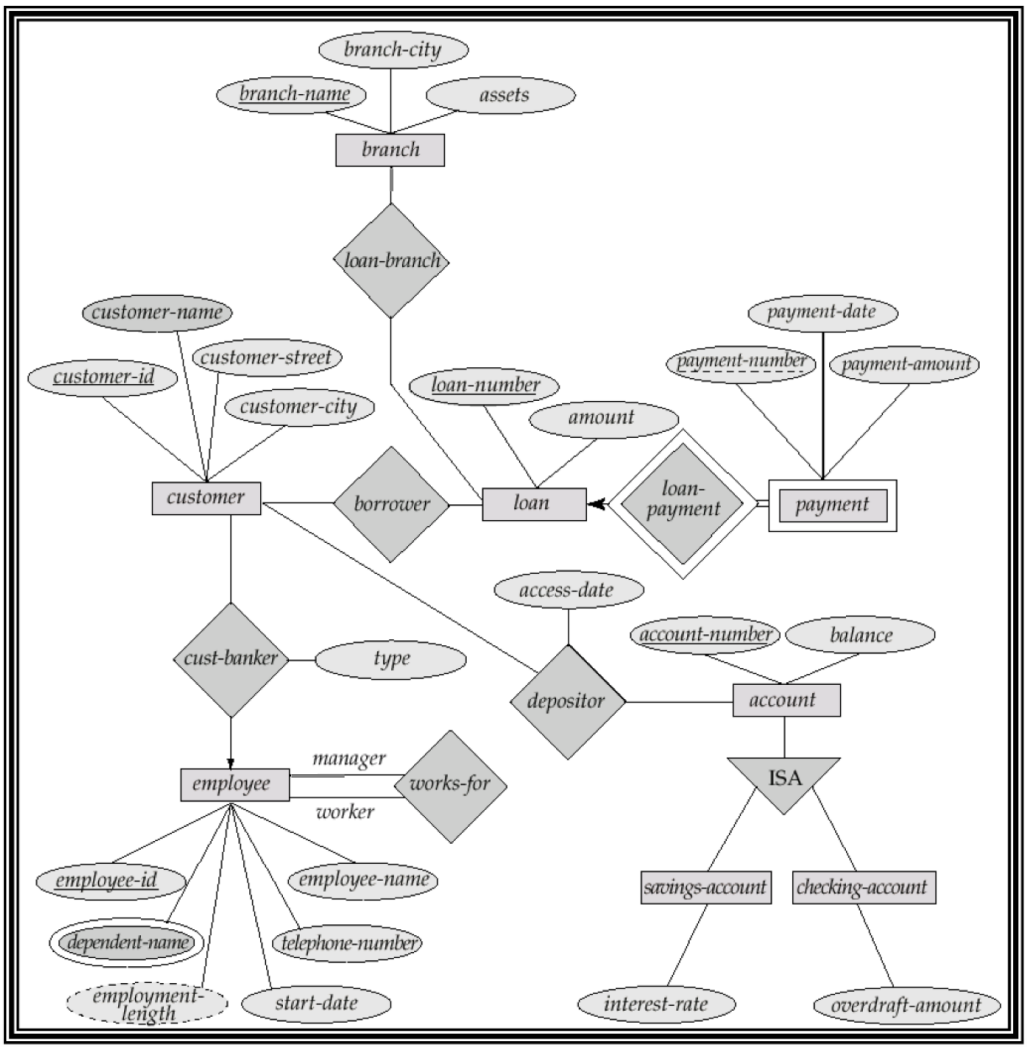
Answer: drives, records, bus in use

(c) Please draw an ER diagram for this database and please indicate keys and participation constraints in the diagram with their corresponding symbols. But you don't need to show the cardinality ratio. (4 marks)

Answer:



Question 2. [Relational Model] (15 marks)



The figure above shows an ER diagram of a banking enterprise. Note that we ignore the cardinality ratio for the simplicity.

1. Please find all strong entity sets in this diagram. (2 marks)

   Answer: Branch, Customer, Loan, Employee, Account, Savings-account, Checking-account

2. Please find all weak entity sets in this diagram. (1 marks)

   Answer: Payment

3. Please find the multivalued attribute and the derived attribute in this diagram. (2 marks)

   Answer: dependent-name, employment-length

4. Please convert each entity set in the diagram into a table. In each table, the key should be underlined. (5 marks)

   Answer:

   Branch (branch-name, branch-city, assets)
   Customer (customer-id, customer-name, customer-street, customer-city)
   Loan (loan-number, amount)
   Employee (employee-id, employee-name, telephone-number, start-date)
   Account (account-number, balance)
   Savings-account (account-number, interest-rate)
   Checking-account (account-number, overdraft-amount)
   Payment (loan-number, payment-number, payment-date, payment-amount)

5. Please convert each relationship set in the diagram into a table. In each table, the key should be underlined. (5 marks)

   Answer:

   Borrower (customer-id, loan-number)
   Depositor (customer-id, account-number, access-date)
   Loan-branch (loan-number, branch-name)
   Cust-banker (customer-id, employee-id, type)
   Works-for (employee-id, manager-id)

Question 3. [SQL] (20 marks)

Consider the following database schema.

- Sailors (sid, sname, rating ,age),
- Reserves (sid, bid, date),
- Boats (bid,bname,color)

1. Please write SQL to find the names of sailors who reserved bid=103 (3 marks)

   Answer:

   ```
   SELECT S.sname
   FROM Sailors as S, Reserves as R
   WHERE S.sid=R.sid AND R.bid=103
   ```

2. Please write SQL to find sid's of sailors who've reserved a red **or** a green boat (4 marks)

   Answer:

   Solution 1:

   ```
   SELECT R.sid
   FROM Boats as B, Reserves as R
   WHERE R.bid=B.bid
   AND (B.color='red' OR B.color='green')
   ```

   Solution 2:

   ```
   (SELECT R.sid
   FROM Boats as B, Reserves as R
   WHERE R.bid=B.bid
   AND B.color='red')
   UNION
   ```

(SELECT R.sid
FROM Boats as B, Reserves as R
WHERE R.bid=B.bid
AND B.color='green')

3. Please write SQL to find sid's of sailors who've reserved a red **and** a green boat (4 marks)

Answer:

Solution 1:

SELECT S.sid
FROM Sailors as S, Boats as B1, Reserves as R1,
Boats as B2, Reserves as R2
WHERE S.sid=R1.sid AND R1.bid=B1.bid
AND B1.color='red' AND S.sid=R2.sid
AND R2.bid=B2.bid AND B2.color='green'

Solution 2:

(SELECT S.sid
FROM Sailors as S, Boats as B, Reserves as R
WHERE S.sid=R.sid AND R.bid=B.bid
AND B.color='red')
INTERSECT
(SELECT S.sid
FROM Sailors as S, Boats as B, Reserves as R
WHERE S.sid=R.sid AND R.bid=B.bid
AND B.color='green')

4. Please write SQL to find the names of sailors who have not reserved bid=103 (3 marks)

Answer:

```
SELECT S.sname
FROM Sailors S
WHERE S.sid NOT IN (SELECT R.sid
FROM Reserves R
WHERE R.bid=103)
```

Find the names of sailors who have not reserved bid=103

5. Please write SQL to find the record of the sailor with the highest rating (3 marks)

Answer:

```
SELECT *
FROM Sailors as S
WHERE S.rating ≥ ALL (SELECT S2.rating
FROM Sailors as S2)
```

6. Please write SQL to find the average age of sailors with rating 10. (3 marks)

Answer:

```
SELECT AVG (S.age)
FROM Sailors as S
WHERE S.rating=10
```

Question 4. [Relational Algebra] (15 marks)

Consider the following database schema.

- Sailors (sid, sname),
- Reserves (sid, bid, date),

- Boats (<u>bid</u>,bname,color)

1. Write relational algebra operations to find the ids of sailors who've reserved boat with bid 103. (3 marks)

   Answer:

   $\pi_{sid}(\sigma_{bid=103}\text{Reserves})$

2. Write relational algebra operations to find the names of sailors who've reserved boat with bid 103. (3 marks)

   Answer:

   Solution 1:

   $\pi_{sname}(\sigma_{\text{Reserves.sid=Sailors.sid} \wedge bid=103}(\text{Reserves x Sailors}))$

   Solution 2:

   $\pi_{sname}(\sigma_{\text{Reserves.sid=Sailors.sid}}((\sigma_{bid=103}\text{Reserves}) \text{ x Sailors}))$

   Solution 3:

   $\pi_{sname}((\sigma_{bid=103}\text{Reserves}) \text{ JOIN}_{sid}\text{Sailors})$

   Note: JOIN operation can be either natural join (*) or equijoin ($\bowtie$).

3. Write relational algebra operations to find the names of sailors who've reserved a red boat. (3 marks)

   Answer:

$\pi_{sname}((\sigma_{color=red}Boats)\ JOIN_{bid}\ Reserves\ JOIN_{sid}Sailors)$

Note: JOIN operation can be either natural join (*) or equijoin ($\bowtie$).

4. Write relational algebra operations to find the names of sailors who've reserved a red or a green boat. (3 marks)

Answer:

$\pi_{sname}((\sigma_{color=red \vee color=green}\ Boats)\ JOIN_{bid}\ Reserves\ JOIN_{sid}Sailors)$

Note: JOIN operation can be either natural join (*) or equijoin ($\bowtie$).

5. Write relational algebra operations to find the names of sailors who've reserved a red and a green boat. (3 marks)

Answer:

π_{sname}

$[\pi_{sid,sname}((\sigma_{color=red}Boats)\ JOIN_{bid}\ Reserves\ JOIN_{sid}Sailors) \cap$

$\pi_{sid,sname}((\sigma_{color=green}Boats)\ JOIN_{bid}\ Reserves\ JOIN_{sid}Sailors)]$

Note: JOIN operation can be either natural join (*) or equijoin ($\bowtie$).

Question 5. [Functional Dependency & Normal Forms] (15 marks)

1. Consider a relation R(X,Y,U,V,W) with the following set of dependencies {{X}→{Y}, {U,V}→{W}, {V}→{X}}. Find the closure of each attribute. (3 marks)
Answer:

X+={X,Y} Y+={Y} U+={U} V+={V,X,Y} W+={W}

2. Consider the relation schema R(A,B,C,D,E) and the set of functional dependencies: F = {{A}→{B}, {A,B}→{C}, {D}→{A,C}}. Find the

closure of each attribute. (3 marks)

Answer:
A+ = {A,B,C} B+ = {B}, C+ = {C} D+ = {D,A,B,C} E+ = {E}

3. Consider the relation schema R(A, B, C, D, E) and the set of functional dependencies F={A→B, BC→E, and ED →A}. List the candidate key(s) of R. Please verify if R is in 3NF. If yes, please elaborate the reason why R is in 3NF. If no, please decompose R in 3NF. (4 marks)

Answer:
Candidate keys of R: ACD, BCD, CDE
R is already in 3NF since all the attributes of R are key attributes.

4. Assume the relation schema R(A,B,C,D,E, F, G, H) and the following set of functional dependencies: AB→E, C→D, D→E, FG→A. Write the candidate key(s) for R and Decompose R in BCNF. (5 marks)

Answer:

Candidate keys of R: BCFGH

Decomposition:

1.AB→E causes a violation in R. Break in R1(ABE), and R2(ABCDFGH)

2.C→D causes a violation in R2. Break R2 in R3(CD), and R4(ABCFGH)

3.D→E causes no violation because E has been removed from R4. However, FG→A causes a violation. Break R4 in R5(FGA) and R6(BCFGH).

Final decomposition: R1(ABE), R3(CD), R5(FGA) and R6(BCFGH)

Question 6. [File Organization & Indexing] (15 marks)
Assume a movie database with the following sizes of each attribute:

- Film (Title : 40 bytes, Director Name: 20 bytes, Year: 4 bytes, Company: 20 bytes)
- Actor (ID: 4 bytes, Name: 20 bytes, Date_of_Birth: 4 bytes)
- There exist 30,000 films in the database and 100,000 actors. Each block has 512 bytes and each pointer is 6 bytes. The blocking factor of a file (bfr) is the number of records that fit in a block.

1. What is the blocking factor bfrF for Film relation and the blocking factor bfrA for Actor relation? (2 marks)

   Answer:

   bfrF $=\lfloor 512/84 \rfloor = 6$, bfrA $= \lfloor 512/28 \rfloor = 18$

2. Assuming that the Film relation is sorted on the Title and there is no index. (4 marks)

   (a) What is the cost (in terms of block accesses) for finding the film with title "Titanic"? (2 marks)

   Answer:

   The file is stored in 30,000/6=5,000 blocks. Cost of finding the film with title "Titanic" with a binary search: $\lceil \log_2 5000 \rceil = 13$ block accesses.

   (b) What is the cost (in terms of block accesses) for finding all the films directed by director "John Woo"? Note that what we consider now is not the average cost but the worst-case cost. (2 marks)

   Answer:

   Cost of finding all the films directed by director "John Woo": We

need sequential scan since sorting is not based on director (5000 pages).

3. Assume that the Actor relation is sorted on the name and you want to create an ordered index on ID (each index entry has the form <ID, pointer>). Note that we need a dense index because sorting is according to name (not ID) and in this dense index, each actor ID has a corresponding entry and there are totally 100,000 entries. (9 marks)

   (a) What is the blocking factor for the index (single-level)? (2 marks)

   Answer:

   bfr_{Aindex}= 512/(4+6)=51

   (b) How many pages are required for these index entries? (2 marks)

   Answer:

   100,000/51=1961

   (c) What is the cost of retrieval based on a single ID using this organization (e.g., "Find actor with ID=100")? (2 marks)

   Answer:

   $\lceil \log_2 5000 \rceil$+1=12.

   (d) If you convert the above index in multiple-level index, how many levels you need (assuming full blocks)? (3 marks)

   Answer:

   At the next level we index 1961 blocks – i.e., index contains 1961/51=39 blocks. We need an additional top level with 1 page

END OF PAPER