CITY UNIVERSITY OF HONG KONG

Course code & title : CS3402 Database Systems

Session : Semester A 2020/21

Time allowed : 1.5 hours

This paper has 8 pages (including this cover page).

1. This paper consists of FIVE questions.

2. Write down your answer in the space provided.

*This is an **open-book** examination.*

Candidates are allowed to use the following materials/aids:

Printed lecture notes, personal notes, and other course handout materials

Materials/aids other than those stated above are not permitted.

No Electronic devices.

Candidates will be subject to disciplinary action if any unauthorized

STUDENT ID	
NAME	

Q1 (20%)	Q2 (20%)	Q3 (20%)	Q4 (22%)	Q5 (18%)	Total (100%)

Problem One: Basic Concepts (20 points)
1) Judge whether each of the following statement is TRUE or FALSE and give your justification for your answer. [15 points]

//Marking tips: Correct answer is worth 2 points; correct explanation 1 point!
(a) Database System is a set of software programs for creating, storing, updating and accessing the data of a database. Oracle, MySQL, Microsoft SQL Server, DB2 and Microsoft Access are all Database Systems.
Answer: False. Database Manage System (DBMS) is a set of software programs for creating, storing, updating and accessing the data of a DB; while Database System is an integrated system of hardware, software, people, procedures, and data.

(b) If a 3NF table has two candidate keys and two candidates have no overlapping attribute, then this table is also in BCNF.

Answer: True. BCNF is different from 3NF only when the table has two or more candidate keys which are overlapping.

(c) The SQL statement SELECT * FROM COURSE_GRADE WHERE GPA > 3.5 OR GPA < 2.0 AND Sex='F' will select the female students with GPA larger than 3.5 or smaller than 2.0.

Answer: False. The priority of logic operator (AND, OR, NOT) is: NOT, AND, OR. And the equivalent statement is SELECT * FROM COURSE_GRADE WHERE GPA >= 3.5 OR (GPA < 2.0 AND Sex='F').

(d) If an attribute set is a superkey, then it must be a candidate key.

Answer: The candidate key is a minimal superkey. Superkey is not necessary to be the candidate key.

(e) Since "view" is the virtual table, if we change the value of tuples in the view, the database state is not changed.

Answer. False. Changes in view will be reflected in the base table, thus the database state is updated.

2 List five major advantages of DBMS compared to traditional file systems. [5 Points]

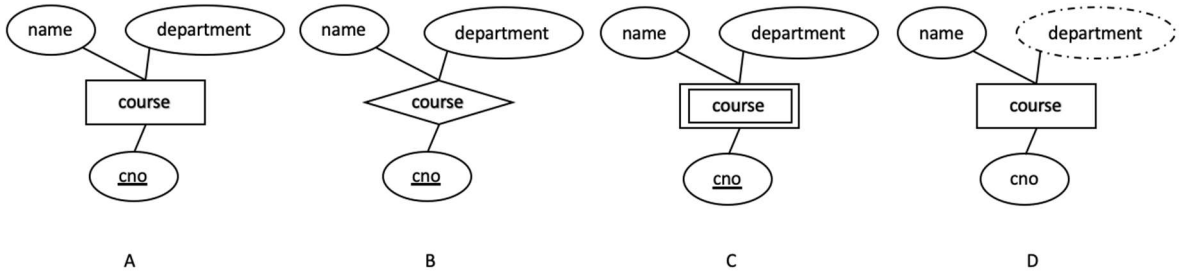
Answer:

Controlling Redundancy, Enforcing Integrity Constraints, Restricting Unauthorized Access, Representing Complex Relationships among Data, Providing Storage Structures and Search Techniques for Efficient Query Processing, Providing Backup and Recovery, Providing Persistent Storage for Program Objects, Providing Multiple User Interfaces and concurrent control, etc
//Marking tips: mention each can get 1 points, total 5 points

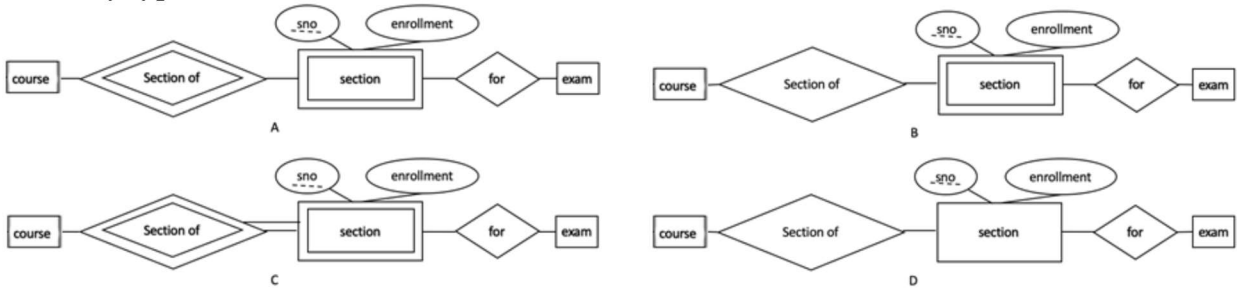
Problem Two: ER Model and Relational Model (20 points)
1 Consider a university database for the scheduling of classrooms for -final exams. This database could be modelled as: 1 There are several course and for each course, it has a course name, a unique course number, and its related department. 2 A course has several sections for the exam, which has a section number and enrolment information. 3 There are also people joining the exams including students and supervisors who has eid and name as their attributes. 4 For each exam, it has a unique exam id and would start at planned time. And the exam would be taken in a room with a unique room number. The exam rooms also have other attributes like capacity and location.

Show an E-R diagram matches the above illustration by selecting possible parts in the following options. [12 points]

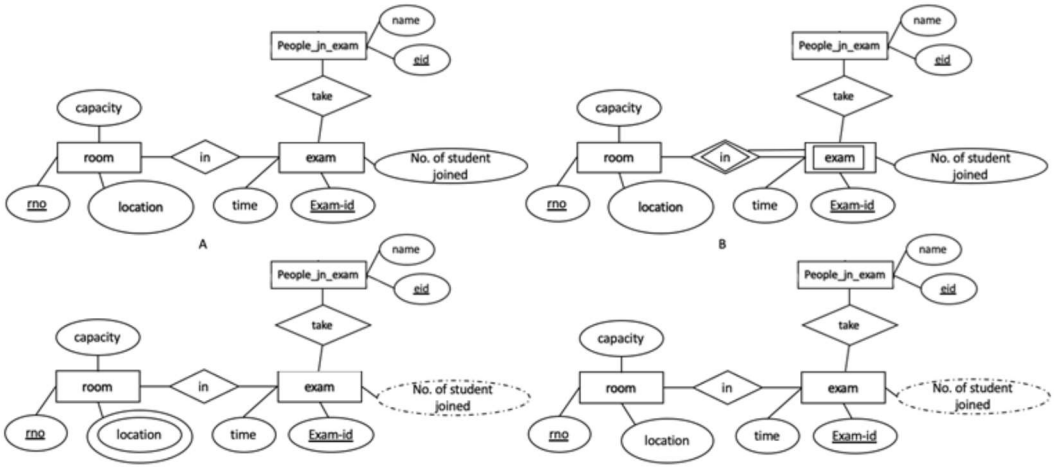
(1) The entity type for course:



(2) The entity type of section



(3) The entity type for people in the exam.



Answer:

ACD

//Marking tips: 4 points for each table

2 Consider the following relational model for a basketball league:
•Player (PlayerID, PlayerName, TeamID)
•Team (TeamID, TeamName, Venue)
•Game (GameNo, HomeTeamID, AwayTeamID)
•Record (GameNo, PlayerID, Points, Rebounds, Assists)
In this basketball league, each team has a unique name and each player plays for only one team. One teamhas at least 10 players. Two teams (home team versus away team) participate in each game at home team's venue. Each team meets all other teams twice (i.e., double round-robin tournament), one time as home team and the other time as away team. For each game, the league records points, the number of rebounds and the number of assists for each player. If a player did not play for a game, there is no record for this player in that game. [8 points]

(a) Suppose Player and Team are two entities and there is a relationship between them. What is the relationship type between Player and Team? Explain your answer. (Relationship types: 1-to-1, 1-to-many, many-to-1 and many-to-many) [4 points]

Answer:

The relationship type from Player to Team is many-to-1 (2marks), because each team has many players and each player can play for only one team (2 marks).

(b) Suppose Player and Team are two entities and there is a relationship between them. What are the participation constraints with the notation (min, max) for (a) the participation of the Player entity in the relationship and (b) the participation of the Team entity in the relationship? [4 points]

Answer:

The participation of the Player entity in the relationship is (1,1). (2marks)

The participation of the Team entity in the relationship is (10, N). (2marks)

Problem Three: Integrity Constraints (20 points)

Suppose we have a relational database containing three tables Employee(emp_id ,name, birth_date, sex, salary, branch_id),Branch(branch_id, branch_name, mgr_id, mgr_start_date) and Client(client_id, client_name, branch_id). The current state of the database is shown in the following tables.

Employee

emp_id	name	birth date	sex	salary	branch_id
100	Davide Wallace	1967-11-17	M	250,000	1
101	Jan Levinson	1961-05-11	F	110,000	1
102	Michael Scott	1964-03-15	M	75,000	2
103	Angela Martin	1971-06-25	F	63,000	2
104	Kelly Kapoor	1980-02-05	F	55,000	2
105	Stanely Hudson	1958-02-19	M	69,000	2
106	Josh Porter	1969-09-05	M	78,000	3
107	Andy Bernard	1973-07-22	M	65,000	3
108	Jim Halpert	1978-10-01	M	71,000	3

Branch

branch_id	branch_name	mgr_id	mgr_start_date
1	Corporate	100	2006-02-09
2	Scranton	102	1992-04-06
3	Stamford	106	1998-02-13

Client

client_id	client_name	branch_id
400	Dunmore Highschool	2
401	Lackawana Country	2
402	FedEx	3
403	John Daly Law, LLC	3

404	Scranton Whitepages	2
405	Times Newspaper	3
406	FedEx	2

For question 2) to 4), Suppose each of the following Update operations is applied directly to the database. Discuss all integrity constraints violated by each operation, if any, and the different ways of enforcing these constraints.

//Marking tips: 2 points for constraints, 3 points for explanation and solutions.

(a) Analyze the primary keys and the foreign keys of providing relations. [5 points]

Answer:
Primary key:
emp_id is the primary key of Employee.
branch_id is the primary key of Branch.
client_id is the primary key of Client.
Foreign key:
The attribute branch_id of relation Employee and relation Client that reference relation Branch.

(b) Insert.<'4','TOG','109','1996-6-5'> into Branch [5 points]

Answer: Violates the referential integrity constraint.
Violates referential integrity because mgr_id='109' and there is no tuple in the Employee relation with emp_id='109'. We may enforce the constraint by: (i) rejecting the insertion, (ii) changing the value of mgr_id to an existing emp_id value in Employee, or (iii) inserting a new Employee tuple with emp_id='109'.

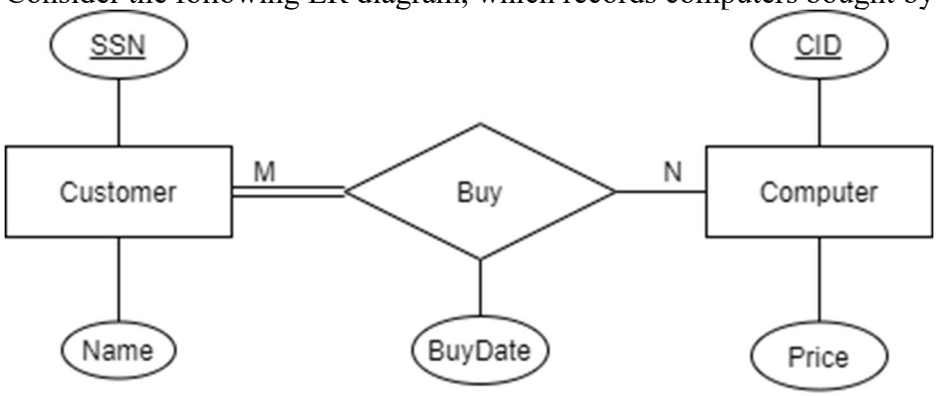
(c) Insert <'103', 'Angela Martin','1964-3-15','F','70,000',null> into Employee [5 points]
Answer: Violates both entity integrity constraint and key constraint.
Violates the entity integrity constraint because branch_id of Branch is the key attribute which does not allow to be NULL. Violates the key constraint because there already exists an Employee tuple with emp_id=103.
We may enforce the constraint by: (i) rejecting the insertion, or (ii) changing the value of branch_id to a value that is not null and already exist in the table Branch. At the same time, we need to change the value of emp_id to a value that doesn't exist in Employee.

(d) Delete the Branch tuple with branch_id=3
Answer: Violates the referential integrity constraint.
Violates the referential integrity constraint because 3 tuples exist in the Employee relations and 3 tuples exist in the Client relations the reference the tuple being deleted from Branch. We may

enforce the constraint by: (i) rejecting the deletion, or (ii) delete all the tuples in the Employee and Client relations whose value for branch_id=3.

Problem Four: SQL (22 points)

1 Consider the following ER diagram, which records computers bought by customers.



Please implement this database by creating tables with SQL statements while maintaining the constraints defined in the ER diagram. SSN and CID are **integers**; Price is the **decimal**. Name is **varying strings less than 32 characters**; BuyDate is the **date**. [6 points]

Answer:
//2points each

```
CREATE TABLE CUSTOMER
(    SSN INTEGER Not NULL,
     Name VARCHAR(32),
     Primary Key (SSN)
);


CREATE TABLE COMPUTER
(    CID INTEGER Not NULL,
     Price FLOAT,
     Primary Key (CID)
);

CREATE TABLE BUYRECORDS
(    SSN INTEGER Not NULL,
     BuyDate DATE,
     CID INTEGER Not NULL,
     Primary Key (SSN, CID),
     Foreign Key (SSN) References CUSTOMER(SSN),
     Foreign Key (CID) References COMPUTER(CID)
);
```

2. Given the following relations about students borrow books from a university library.

Student (**StudentID**: *integer*, **Name**: *string*, **EnrollDate**: *date*)
Book(**Title**: *string*, **Author**: *string*, **Quantity**: *integer*)
BorrowRecord(**StudentID**: *integer*, **BookTitle**: *string*, **BookAuthor**: *string*, **BorrowDate**: *date*, **DueDate**: *date,* **Returned**: *integer*)

Note that there are no duplicate records in the three relations, and there can be duplicates of book titles. For **Returned**, 0 means not returned and 1 indicates the books have been returned.
Write the following queries in SQL.

(a) Find the **StudentID**s of students who don't return the books on time (Assume today is "23-OCT-20"). [4 points]

Answer:

```
SELECT StudentID from BorrowRecord WHERE DueDate < '23-OCT-20'
AND Returned=0;
```

(b) Find the **Name**s of students who have borrowed the book "*SQL theory*" written by "Paul". [4 points]

Answer:

```
SELECT Name FROM Student S JOIN BorrowRecord BR ON S.StudentID =
BR.StudentID WHERE BookAuthor='Paul' and BookTitle='SQL theory';
```

(c) Find the **Title**s and corresponding **Author**s of books which have never been borrowed. [4 points]

Answer:

```
SELECT Title, Author FROM Book WHERE NOT EXISTS (SELECT *
FROM BorrowRecord WHERE Title=BookTitle AND Author=BookAuthor);
```

(d) Find the **Name**s of students who have borrowed every book written by 'Paul'[4 points]

Answer:

```
SELECT Name FROM Student S WHERE NOT EXISTS (SELECT * FROM
Book WHERE Author = 'Paul' AND NOT EXISTS (SELECT * FROM
BorrowRecord BR WHERE BookAuthor=Author AND BookTitle=Title AND
S.StudentID=BR.StudentID));
```

Problem Five: Normalization [18 points]

Let's consider the following table storing the information about various dealers for various products. And it stores the dealer details along with the products we purchase from them and the product count.

Dealer_ID	Dealer_Name	Product_ID	Quantity
D102	Jade Garden	P121	66
D102	Jade Garden	P103	188
D110	Tim Ho Wan	P102	88
D105	Goose Manor	P121	66
D104	Daikiya	P110	288
D104	Daikiya	P102	10

(1) Identify all Functional Dependencies for this table. [4 points]

Answer:
Dealer_ID -> Dealer_Name
Dealer_Name -> Dealer_ID
Dealer_ID Product_ID -> Dealer_Name Quantity
Dealer_Name Product_ID -> Dealer_ID Quantity

(2) List all candidate keys of this table. [4 points]
Answer:
(Dealer_ID, Product_ID)
(Dealer_Name, Product_ID)

(3) Is the relation in 3NF and why? If not, normalize it into 3NF . [5 points]
Answer:
Yes, because, no non-prime Transitive Functional Dependency on candidate key is held.

(4) Is the relation in BCNF and why? If not, normalize it into BCNF . [5 points]
Answer:
No, because Dealer_ID -> Dealer_Name, Dealer_ID is not the superkey, same for Dealer_Name -> Dealer_ID.

The result of decomposition:
T1(Dealer_ID, Dealer_Name)
T2 (Dealer_ID, Product_ID, Quantity)