Question A. (50 marks)

1) A database is being constructed to keep track of the teams and games of a football league:

*the league has many teams, each team has a name, a city, a coach, a captain, and a set of players.

*each player belongs to only one team, each player has an id, a name, a position (such as right back or center forward), a skill level, and a set of injury records.

* a team captain is also a player

* each injury record keeps track of the player who has been treated, the date of the treatment, several drugs used in the treatment.

* a game is played between two teams (referred to as host_team and guest_team) and has a date (such as June 11th, 2015) and a score (such as 4 to 2).

Design an ER diagram for this application, stating any assumptions you make. [30 marks]

2) For your ER diagram given above, convert it into relational schema using the mapping guidelines discussed in the lecture. For each relation (table) obtained, specify the name and its attributes, as well as its primary key. [20 marks]

**Question B. [50 marks]**

Consider the relation R = {A, B, C, D, E, F, G, H, I,J} and the set of functional dependencies F = { {A, B} -> {C}, {A} -> {D, E}, {B} -> {F}, {F} ->{G, H}, {D} -> {I, J} }.

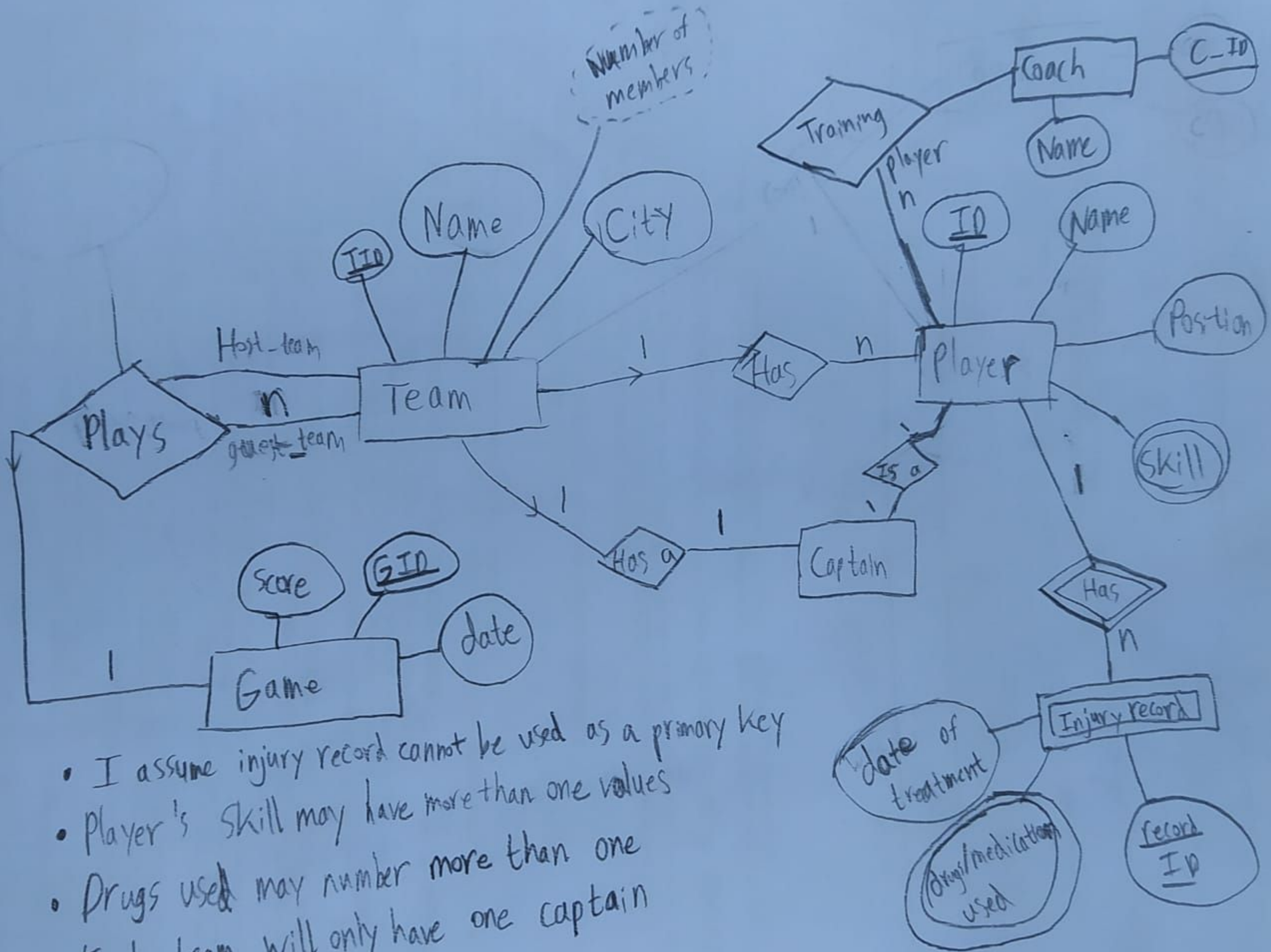
(a) Proof {A}->{E,J} holds by using inference rules. (10 marks)

(b) Whether {A,B,C} is a super key? Whether {A,B,C} is a candidate key? Why? (10 marks)

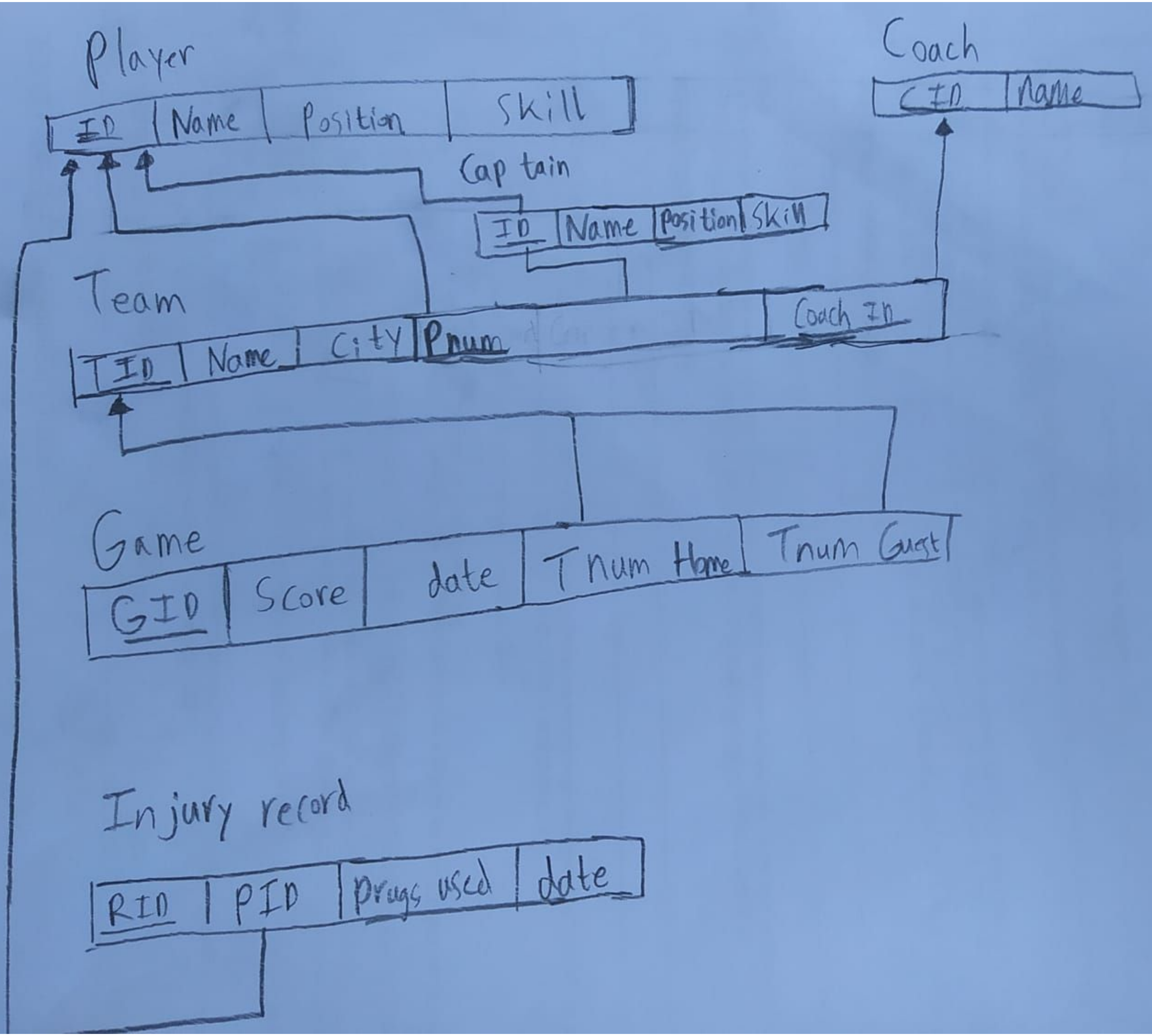
(c) Whether {A,F} is a super key? Whether {A,F} is a candidate key? Why? (10 marks)

(d) Decompose R into 2NF. (10 marks)

(e) Decompose R into BCNF. (10 marks)




- I assume injury record cannot be used as a primary key
- Player's skill may have more than one values
- Drugs used may number more than one
- Each team will only have one captain
- The coach will obviously train the team


Player
ID
Name
Position
Skill
Coach
CID
Name
Captain
ID
Name
Position
Skill
Team
TID
Name
City
Pnum
Coach ID
Game
GID
Score
date
Tnum Home
Tnum Guest
Injury record
RID
PID
Drugs used
date

$R = \{A, B, C, D, E, F, G, H, I, J\}$

$= \{ \{A, B\} \rightarrow \{C\}, \{A\} \rightarrow \{D, E\}, \{B\} \rightarrow \{F\}, \{F\} \rightarrow \{G, H\},$
$\{D\} \rightarrow \{I, J\} \}$

a) $A^+ = C\,D\,E\,I\,J$ by reflexive

$\hookrightarrow D^+ = \cancel{D}\,I\,J$ by transitive

$\{A\} \rightarrow \{E, J\}$ works

b) $A^+ = \cancel{A}C\,D\,E\,I\,J$ , not a super

$B^+ = \cancel{B}\,C\,F\,G\,H$ , Not a super

$\hookrightarrow F^+ = \cancel{F}\,G\,H$

$C^+ = \cancel{C}$ Not a super

$ABC^+ = ABCDEIJ$ , Is a super

$\therefore \{ABC\}$ is a candidate key , not a superkey

c) $A^+ = \cancel{A}CDEIJ$, Not a super

$F^+ = \cancel{F}GH$ , Not a super

$\cancel{A}F^+ = ACDEFGHIJ$ , Not a super

$\therefore \{A, F\}$ Not a super and not candidate key

d) $R_1 = \{A, B, C\}$

$A^+ = A\cancel{CDEIJ}$

$B^+ = C\cancel{FGH}$

$C^+ = C$

$D^+ = IJ$

$F^+ = \{GH\}$

$R_2 = \{A, D, E, I, J\}$

$R_3 = \{B, F, G, H\}$

$\therefore R_1, R_2, R_3$, only has One superkey each and no partial dependencies

e) $R_{2a} = \{ B, F \}$ $B \to F \to GH$

$R_{2b} = \{ F, G, H \}$

$R_{3a} = \{ A, D, E \}$ $A \to D \to IJ$

$R_{3b} = \{ D, I, J \}$ $E \to \emptyset$

# CS3402 Database Systems

# Homework

Question A. (50 marks)

A database schema consisting of three relations STUDENT, COURSE, and STAFF is created as follows:

```
CREATE TABLE STUDENT (STU_ID CHAR(4),
                      STUDENT_NAME CHAR(20),
                      ADDRESS CHAR(20),
                      BIRTHDATE DATE,
                      GENDER CHAR(6));

CREATE TABLE COURSE (COURSE_ID CHAR(6),
                     COURSE_TITLE CHAR(20),
                     STAFF_ID CHAR(3),
                     SECTION NUMBER(2));

CREATE TABLE STAFF (STAFF_ID CHAR(3),
                    STAFF_NAME CHAR(20),
                    GENDER CHAR(6),
                    DEPARTMENT CHAR(20),
                    BOSS_ID CHAR(3)
                    SALARY NUMBER(8,2));
```

Write down a SQL statement for each query below:

1) List the names of all male students who were born before 01-01-1995. [5 marks]

2) List the names of all students whose name is at least 6 characters long and whose birthdate falls within 01-01-1995 and 01-01-2000. Order the results alphabetically. [5 marks]

3) List the names of all courses that are neither taught by SMITH nor by JONES. [6 marks]

4) Retrieve the number of courses for each section number. Your output should be in ascending order of section numbers. [6 marks]

5) Find the name of every staff member who teaches the exact same number of courses as his/her boss. [7 marks]

6) List all staff members whose salary is higher than the average salary. [7 marks]

7) Find the names of the staff members who are the boss of some staff member in the Accounting department. [7 marks]

8) Find the average number of courses taught per staff member. [7 marks]

Question B. [50 marks]

Specify the following queries on the COMPANY relational database schema shown in Figure 5.5 below using relational algebra expressions. Also **show the result** of each query as it would apply to the database state of Figure 5.6.

**EMPLOYEE**

| Fname | Minit | Lname | Ssn | Bdate | Address | Sex | Salary | Super_ssn | Dno |
|---|---|---|---|---|---|---|---|---|---|

**DEPARTMENT**

| Dname | Dnumber | Mgr_ssn | Mgr_start_date |
|---|---|---|---|

**DEPT_LOCATIONS**

| Dnumber | Dlocation |
|---|---|

**PROJECT**

| Pname | Pnumber | Plocation | Dnum |
|---|---|---|---|

**WORKS_ON**

| Essn | Pno | Hours |
|---|---|---|

**DEPENDENT**

| Essn | Dependent_name | Sex | Bdate | Relationship |
|---|---|---|---|---|

**Figure 5.5**
Schema diagram for the COMPANY relational database schema.

**Figure 5.6**
One possible database state for the COMPANY relational database schema.

**EMPLOYEE**

| Fname | Minit | Lname | Ssn | Bdate | Address | Sex | Salary | Super_ssn | Dno |
|---|---|---|---|---|---|---|---|---|---|
| John | B | Smith | 123456789 | 1965-01-09 | 731 Fondren, Houston, TX | M | 30000 | 333445555 | 5 |
| Franklin | T | Wong | 333445555 | 1955-12-08 | 638 Voss, Houston, TX | M | 40000 | 888665555 | 5 |
| Alicia | J | Zelaya | 999887777 | 1968-01-19 | 3321 Castle, Spring, TX | F | 25000 | 987654321 | 4 |
| Jennifer | S | Wallace | 987654321 | 1941-06-20 | 291 Berry, Bellaire, TX | F | 43000 | 888665555 | 4 |
| Ramesh | K | Narayan | 666884444 | 1962-09-15 | 975 Fire Oak, Humble, TX | M | 38000 | 333445555 | 5 |
| Joyce | A | English | 453453453 | 1972-07-31 | 5631 Rice, Houston, TX | F | 25000 | 333445555 | 5 |
| Ahmad | V | Jabbar | 987987987 | 1969-03-29 | 980 Dallas, Houston, TX | M | 25000 | 987654321 | 4 |
| James | E | Borg | 888665555 | 1937-11-10 | 450 Stone, Houston, TX | M | 55000 | NULL | 1 |

**DEPARTMENT**

| Dname | Dnumber | Mgr_ssn | Mgr_start_date |
|---|---|---|---|
| Research | 5 | 333445555 | 1988-05-22 |
| Administration | 4 | 987654321 | 1995-01-01 |
| Headquarters | 1 | 888665555 | 1981-06-19 |

**DEPT_LOCATIONS**

| Dnumber | Dlocation |
|---|---|
| 1 | Houston |
| 4 | Stafford |
| 5 | Bellaire |
| 5 | Sugarland |
| 5 | Houston |

**WORKS_ON**

| Essn | Pno | Hours |
|---|---|---|
| 123456789 | 1 | 32.5 |
| 123456789 | 2 | 7.5 |
| 666884444 | 3 | 40.0 |
| 453453453 | 1 | 20.0 |
| 453453453 | 2 | 20.0 |
| 333445555 | 2 | 10.0 |
| 333445555 | 3 | 10.0 |
| 333445555 | 10 | 10.0 |
| 333445555 | 20 | 10.0 |
| 999887777 | 30 | 30.0 |
| 999887777 | 10 | 10.0 |
| 987987987 | 10 | 35.0 |
| 987987987 | 30 | 5.0 |
| 987654321 | 30 | 20.0 |
| 987654321 | 20 | 15.0 |
| 888665555 | 20 | NULL |

**PROJECT**

| Pname | Pnumber | Plocation | Dnum |
|---|---|---|---|
| ProductX | 1 | Bellaire | 5 |
| ProductY | 2 | Sugarland | 5 |
| ProductZ | 3 | Houston | 5 |
| Computerization | 10 | Stafford | 4 |
| Reorganization | 20 | Houston | 1 |
| Newbenefits | 30 | Stafford | 4 |

**DEPENDENT**

| Essn | Dependent_name | Sex | Bdate | Relationship |
|---|---|---|---|---|
| 333445555 | Alice | F | 1986-04-05 | Daughter |
| 333445555 | Theodore | M | 1983-10-25 | Son |
| 333445555 | Joy | F | 1958-05-03 | Spouse |
| 987654321 | Abner | M | 1942-02-28 | Spouse |
| 123456789 | Michael | M | 1988-01-04 | Son |
| 123456789 | Alice | F | 1988-12-30 | Daughter |
| 123456789 | Elizabeth | F | 1967-05-05 | Spouse |

(a) Retrieve the names of employees that have at least one dependent. [5 marks]

(b) List the first names of employees who work in the Research department and whose salary is below 39000. [5 marks]

(c) Find the first names of male employees that are not supervised by ‘Jennifer Wallace'. [5 marks]

(d) List the first names of all employees who work on the project ‘Newbenefits’ and who have a daughter as a dependent. [5 marks]

(e) Retrieve the first names of employees who work on every project. [5 marks]

(f) Retrieve the first names of employees who do not work on any project. [7.5 marks]

(g) Retrieve the salary of all male employees working in the department with department number 4. [5 marks]

(h) Find the first names and addresses of employees who work on at least one project located in Houston but whose department has its location in either Sugarland or Bellaire. [7.5 marks]

(i) List the last name of each department manager who has a spouse as a dependent. [5 marks]

# CS3402 Database Systems

# Homework

Question A. (50 marks)

A database schema consisting of three relations STUDENT, COURSE, and STAFF is created as follows:

```
CREATE TABLE STUDENT (STU_ID CHAR(4),
                      STUDENT_NAME CHAR(20),
                      ADDRESS CHAR(20),
                      BIRTHDATE DATE,
                      GENDER CHAR(6));

CREATE TABLE COURSE (COURSE_ID CHAR(6),
                     COURSE_TITLE CHAR(20),
                     STAFF_ID CHAR(3),
                     SECTION NUMBER(2));

CREATE TABLE STAFF (STAFF_ID CHAR(3),
                    STAFF_NAME CHAR(20),
                    GENDER CHAR(6),
                    DEPARTMENT CHAR(20),
                    BOSS_ID CHAR(3)
                    SALARY NUMBER(8,2));
```

Write down a SQL statement for each query below:

1) List the names of all male students who were born before 01-01-1995. [5 marks]

```
select student_name from student where gender='m' and birthdate < '01-JAN-1995';
```

*Note: using the date format 01-01-1995 is also fine.*

2) List the names of all students whose name is at least 6 characters long and whose birthdate falls within 01-01-1995 and 01-01-2000. Order the results alphabetically. [5 marks]

```
SELECT student_name FROM student WHERE length(student_name) >= 6 AND
birthdate BETWEEN '01-JAN-1995' and '01-JAN-2000'
```

3) List the names of all courses that are neither taught by SMITH nor by JONES.
[6 marks]

```
SELECT course_title from COURSE
WHERE STAFF_ID NOT IN
(SELECT staff_ID
  FROM staff
  WHERE staff_name IN ('JONES','SMITH'))
```

4) Retrieve the number of courses for each section number. Your output should be in ascending order of section numbers. [6 marks]

```
SELECT section, COUNT(*)
FROM course
GROUP BY section
ORDER BY section
```

5) Find the name of every staff member who teaches the exact same number of courses as his/her boss. [7 marks]

```
SELECT staff_name
FROM STAFF S1
WHERE
(SELECT COUNT (*) FROM COURSE WHERE STAFF_ID=S1.STAFF_ID)
=
(SELECT COUNT (*) FROM COURSE WHERE STAFF_ID=S1.BOSS_ID);
```

6) List all staff members whose salary is higher than the average salary. [7 marks]

```
SELECT *
FROM staff
WHERE salary > (SELECT AVG(SALARY) FROM staff);
```

7) Find the names of the staff members who are the boss of some staff member in the Accounting department. [7 marks]

```
SELECT staff_name
FROM staff
WHERE staff_id IN
(SELECT boss_id
 FROM staff
 WHERE department='Accounting');
```

8) Find the average number of courses taught per staff member. [7 marks]

```
SELECT COUNT(DISTINCT course_id) / COUNT(DISTINCT staff_id)
FROM course;
```

*Note: It is fine to write COUNT(course_id), i.e. without 'DISTINCT' for the first count expression assuming that course_id is a key attribute of course. Also, it is fine to divide by the total number of staff members (i.e. even the ones that do not teach any courses.)*

Question B. [50 marks]

Specify the following queries on the COMPANY relational database schema shown in Figure 5.5 below using relational algebra expressions. Also show the result of each query as it would apply to the database state of Figure 5.6. (Note that some of the queries may return an empty result set.)

**EMPLOYEE**

| Fname | Minit | Lname | Ssn | Bdate | Address | Sex | Salary | Super_ssn | Dno |
|---|---|---|---|---|---|---|---|---|---|

**DEPARTMENT**

| Dname | Dnumber | Mgr_ssn | Mgr_start_date |
|---|---|---|---|

**DEPT_LOCATIONS**

| Dnumber | Dlocation |
|---|---|

**PROJECT**

| Pname | Pnumber | Plocation | Dnum |
|---|---|---|---|

**WORKS_ON**

| Essn | Pno | Hours |
|---|---|---|

**DEPENDENT**

| Essn | Dependent_name | Sex | Bdate | Relationship |
|---|---|---|---|---|

**Figure 5.5**
Schema diagram for the COMPANY relational database schema.

**Figure 5.6**
One possible database state for the COMPANY relational database schema.

**EMPLOYEE**

| Fname | Minit | Lname | Ssn | Bdate | Address | Sex | Salary | Super_ssn | Dno |
|---|---|---|---|---|---|---|---|---|---|
| John | B | Smith | 123456789 | 1965-01-09 | 731 Fondren, Houston, TX | M | 30000 | 333445555 | 5 |
| Franklin | T | Wong | 333445555 | 1955-12-08 | 638 Voss, Houston, TX | M | 40000 | 888665555 | 5 |
| Alicia | J | Zelaya | 999887777 | 1968-01-19 | 3321 Castle, Spring, TX | F | 25000 | 987654321 | 4 |
| Jennifer | S | Wallace | 987654321 | 1941-06-20 | 291 Berry, Bellaire, TX | F | 43000 | 888665555 | 4 |
| Ramesh | K | Narayan | 666884444 | 1962-09-15 | 975 Fire Oak, Humble, TX | M | 38000 | 333445555 | 5 |
| Joyce | A | English | 453453453 | 1972-07-31 | 5631 Rice, Houston, TX | F | 25000 | 333445555 | 5 |
| Ahmad | V | Jabbar | 987987987 | 1969-03-29 | 980 Dallas, Houston, TX | M | 25000 | 987654321 | 4 |
| James | E | Borg | 888665555 | 1937-11-10 | 450 Stone, Houston, TX | M | 55000 | NULL | 1 |

**DEPARTMENT**

| Dname | Dnumber | Mgr_ssn | Mgr_start_date |
|---|---|---|---|
| Research | 5 | 333445555 | 1988-05-22 |
| Administration | 4 | 987654321 | 1995-01-01 |
| Headquarters | 1 | 888665555 | 1981-06-19 |

**DEPT_LOCATIONS**

| Dnumber | Dlocation |
|---|---|
| 1 | Houston |
| 4 | Stafford |
| 5 | Bellaire |
| 5 | Sugarland |
| 5 | Houston |

**WORKS_ON**

| Essn | Pno | Hours |
|---|---|---|
| 123456789 | 1 | 32.5 |
| 123456789 | 2 | 7.5 |
| 666884444 | 3 | 40.0 |
| 453453453 | 1 | 20.0 |
| 453453453 | 2 | 20.0 |
| 333445555 | 2 | 10.0 |
| 333445555 | 3 | 10.0 |
| 333445555 | 10 | 10.0 |
| 333445555 | 20 | 10.0 |
| 999887777 | 30 | 30.0 |
| 999887777 | 10 | 10.0 |
| 987987987 | 10 | 35.0 |
| 987987987 | 30 | 5.0 |
| 987654321 | 30 | 20.0 |
| 987654321 | 20 | 15.0 |
| 888665555 | 20 | NULL |

**PROJECT**

| Pname | Pnumber | Plocation | Dnum |
|---|---|---|---|
| ProductX | 1 | Bellaire | 5 |
| ProductY | 2 | Sugarland | 5 |
| ProductZ | 3 | Houston | 5 |
| Computerization | 10 | Stafford | 4 |
| Reorganization | 20 | Houston | 1 |
| Newbenefits | 30 | Stafford | 4 |

**DEPENDENT**

| Essn | Dependent_name | Sex | Bdate | Relationship |
|---|---|---|---|---|
| 333445555 | Alice | F | 1986-04-05 | Daughter |
| 333445555 | Theodore | M | 1983-10-25 | Son |
| 333445555 | Joy | F | 1958-05-03 | Spouse |
| 987654321 | Abner | M | 1942-02-28 | Spouse |
| 123456789 | Michael | M | 1988-01-04 | Son |
| 123456789 | Alice | F | 1988-12-30 | Daughter |
| 123456789 | Elizabeth | F | 1967-05-05 | Spouse |

(a) Retrieve the names of employees that have at least one dependent. [5 marks]

$\pi_{\text{Fname,Minit,Lname}}(\text{EMPLOYEES} \bowtie_{\text{Essn = Ssn}} \text{DEPENDENT})$

Result: John B Smith, Franklin T Wong, Jennifer S Wallace

(b) List the first names of employees who work in the Research department and whose salary is below 39000. [5 marks]

$\pi_{\text{Fname}}(\sigma_{\text{salary<39000 AND Dname='Research'}}\ \text{EMPLOYEE} \bowtie_{\text{Dno=Dnumber}}\ \text{DEPARTMENT})$

Results: John, Ramesh, Joyce

(c) Find the first names of male employees that are not supervised by 'Jennifer Wallace'. [5 marks]

$Jennifer \leftarrow \sigma_{Fname='Jennifer'\ AND\ Lname = 'Wallace'}(EMPLOYEE)$
$Jennifer\_emp \leftarrow \pi_{EMPLOYEE.Fname}(EMPLOYEE \bowtie_{Super_{ssn}=Ssn} Jennifer)$
$Male\_emp \leftarrow \pi_{Fname}(\sigma_{Sex='M'}(EMPLOYEE))$
Result $\leftarrow Male\_emp - Jennifer\_emp$

Results: John, Franklin, Ramesh, James

(d) List the first names of all employees who work on the project 'Newbenefits' and who have a daughter as a dependent. [5 marks]

Relation $\leftarrow \sigma_{\text{relationship='Daughter'}}(\text{EMPLOYEE} \bowtie_{\text{Essn=Ssn}} \text{DEPENDENT})$
Employees $\leftarrow \sigma_{\text{Pname='Newbenefits'}}(\text{EMPLOYEE} \bowtie_{\text{Dno=Dnum}} \text{PROJECT})$
$\pi_{\text{Fname}}(\text{Relation} \cap \text{Employees})$

Results: (empty)

(e) Retrieve the first names of employees who work on every project. [5 marks]

Ssn_pno $\leftarrow \rho_{(\text{Ssn, Pnumber})}(\pi_{\text{Essn,Pno}}((\text{WORKS\_ON}))$

PNos $\leftarrow \pi_{\text{Pnumber}}(\text{PROJECT})$

SSn_all_pnos $\leftarrow$ SSn_pno ÷ PNos

$\pi_{\text{Fname}}(\text{EMPLOYEE} * \text{SSn\_all\_pnos})$

Result: (empty)

(f) Retrieve the first names of employees who do not work on any project. [7.5 marks]

ALL_EMPS <-- $\pi_{\text{Ssn}}(\text{EMPLOYEE})$

WORKING_EMPS<-- $\rho_{\text{Ssn}}(\pi_{\text{Essn}}(\text{WORKS\_ON}))$

NON_WORKING_EMPS <-- ALL_EMPS - WORKING_EMPS

$\pi_{\text{Fname}}(\text{EMPLOYEE} * \text{NON\_WORKING\_EMPS})$

Result: (empty)

(g) Retrieve the salary of all male employees working in the department with department number 4. [5 marks]

$\pi_{\text{salary}}(\sigma_{\text{sex='M' and DNo=4}}(\text{EMPLOYEE}))$

Results: 25000

(h) Find the first names and addresses of employees who work on at least one project located in Houston but whose department has its location in either Sugarland or Bellaire. [7.5 marks]

Houston_projects $\leftarrow \sigma_{\text{PLocation='Houston'}}$ (PROJECTS $\bowtie_{\text{PNumber=PNo}}$ WORKS_ON)

Dptmts $\leftarrow \sigma_{\text{DLocation='Sugarland' or DLocation='Bellaire'}}$ (DEPARTMENT * DEPT_LOCATION)

Employee_Dptmts $\leftarrow$ Dptmts $\bowtie_{\text{DNo=DNumber}}$ EMPLOYEE

$\pi_{\text{Fname,Address}}$ (Houston_projects $\bowtie_{\text{Essn =Ssn}}$ Employee_Dptmts)

Results:
Ramesh, 975 Fire Oak, Humble TX
Franklin, 638 Voss, Houston TX

(i) List the last name of each department manager who has a spouse as a dependent. [5 marks]

Mgr_ssn_spouse $\leftarrow \sigma_{\text{relationship='Spouse'}}$ (DEPARTMENT $\bowtie_{\text{Mgr\_ssn=Ssn}}$ DEPENDENT)

$\pi_{\text{Lname}}$ (Mgr_ssn_spouse $\bowtie_{\text{Mgr\_ssn=Essn}}$ EMPLOYEE)

Results: Wallace, Wong

**Question 1. (50 marks)**

Suppose that you have formatted your disk with a block size of 2048 bytes and assume that we have 50,000 CAR records of fixed length. A block pointer is 5 bytes long (P=5), and a record pointer is 6 bytes long (Pr=6). Each CAR record has the following fields: Model (20 bytes), Registration (8 bytes), Color (9 bytes), License (9 bytes), Location (40 bytes), Mileage (10 bytes), Price (8 bytes), Year (4 bytes), Manufacturer (16 bytes), and Remarks (200 bytes). The file is ordered by the key field Registration and we want to construct a primary index on Registration.

(a) Calculate the blocking factor (bfr) and the number of file blocks needed to store the CAR records. Assume that records are stored unspanned. How much space remains unused per block? . [10 marks]

(b) Calculate the index blocking factor bfri . [10 marks]

(c) Assume that the primary index is a single-level index. Calculate the number of index entries and the number of index blocks. [10 marks]

(d) Now suppose that we want to make the primary index a multilevel index. How many levels are needed and what is the total number of blocks required by the multilevel index? . [10 marks]

(e) Consider the multilevel index from subquestion (d). What is the number of block accesses needed to search for and retrieve a record from the file given its Registration value? . [10 marks]

**Question 2. [20 marks]**

Suppose that the size of a search key field is V=9 bytes, the size of a record pointer is Pr=7 bytes, the size of a block pointer/tree pointer is P=6 bytes, the order of internal nodes (p) is 20, and the number of data record pointers for a leaf node is 20 for a B+ tree. What should be the minimum required block size B for this B+ tree.

**Question 3. [30 marks]**

Consider three transactions T1, T2, and T3 and two schedules S1, S2, which are given below :

T1: r1 (X); r1 (Z); w1 (X);

T2: r2 (Z); r2 (Y); w2 (Z); w2 (Y);

T3: r3 (X); r3 (Y); w3 (Y);

S1: r1(X); r2(Z); r1 (Z); r3(X); r3(Y); w1(X); c1; w3(Y); r2(Y); c3; w2(Z); w2(Y); c2;

S2: r1(X); r2 (Z); r3 (X); r1(Z); r2 (Y); r3 (Y); w1 (X); c1; w2 (Z); w3 (Y); w2 (Y); c3; c2;

1 ) Draw the serialization graphs for S1, S2 and state whether each schedule is serializable or not. If a schedule is serializable, write down all equivalent serial schedule(s). [20 marks]

2) Determine the strictest recoverability condition (non-recoverable, recoverable, cascadeless or strict) that each schedule satisfies.[10 marks]

# CS3402 Database Systems

## Homework 3

Question 1. (50 marks)

Suppose that you have formatted your disk with a block size of 2048 bytes and assume that we have 50,000 CAR records of fixed length. A block pointer is 5 bytes long (P=5), and a record pointer is 6 bytes long (Pr=6). Each CAR record has the following fields: Model (20 bytes), Registration (8 bytes), Color (9 bytes), License (9 bytes), Location (40 bytes), Mileage (10 bytes), Price (8 bytes), Year (4 bytes), Manufacturer (16 bytes), and Remarks (200 bytes). The file is ordered by the key field Registration and we want to construct a primary index on Registration.

(a) Calculate the blocking factor (bfr) and the number of file blocks needed to store the CAR records. Assume that records are stored unspanned. How much space remains unused per block? . [10 marks]

Solution: The record size R = 20 + 8 + 9 + 9 + 40 + 10 + 8 + 4 + 16 + 200 = 324. To calculate the blocking factor, we get bfr = floor( B / R ) = 6 records/block. The number of blocks is I = ceiling(50,000 / bfr) = 8334 blocks. Therefore, the unused space per block is B – (R * bfr) = 104 bytes.

(b) Calculate the index blocking factor bfri . [10 marks]

Solution: Each index entry has size Ri = size(registration) + P = 8 + 5 = 13 bytes. The index blocking factor is bfri = floor( B / Ri) = 157 entries / block.

(c) Assume that the primary index is a single-level index. Calculate the number of index entries and the number of index blocks. [10 marks]

Solution: We need one index entry per file block, so to calculate the number of index and we already calculate the number of blocks needed to store all 50,000 file records in Q1, which is I=8334. Thus, the number of index blocks: Bi = ceil( I / bfri) = 54 blocks.

(d) Now suppose that we want to make the primary index a multilevel index. How many levels are needed and what is the total number of blocks required by the multilevel index? . [10 marks]

Solution: The fan-out for the multilevel index is the same as the index blocking factor (bfri) which is 157 (see above). The number of 1st level blocks L1 is already calculated in 2b, and so we have L1=54. The number of 2nd level blocks is $L2 = ceil(L1 / fo) = 1$. The total number of index blocks is $L1 + L2 = 54 + 1 = 55$ blocks.

(e) Consider the multilevel index from subquestion (d). What is the number of block accesses needed to search for and retrieve a record from the file given its Registration value? . [10 marks]

Solution: This is given as the number of index levels + 1, which, according to c) is 2+1=3.

Question 2. [20 marks]
Suppose that the size of a search key field is V=9 bytes, the size of a record pointer is Pr=7 bytes, the size of a block pointer/tree pointer is P=6 bytes, the order of internal nodes (p) is 20, and the number of data record pointers for a leaf node 20 for a B+ tree. What should be the minimum required block size B for this B+ tree.

Solution:
For internal nodes:
$(p * P) + ((p - 1) * V) \leq B$
$(20 * 6) + ((20 - 1) * 9) \leq B$
$291 \leq B$ [10 marks]

For leaf nodes:
$(pleaf * (Pr + V)) + P \leq$
$B$ $(20 * (7+ 9)) + 6 \leq B$
$326 \leq B$ [10 marks]

The minimum required block size B is 326.

Question 3. [30 marks]

Consider three transactions T1, T2, and T3 and two schedules S1, S2, which are given below:

T1: r1 (X); r1 (Z); w1 (X);

T2: r2 (Z); r2 (Y); w2 (Z); w2 (Y);

T3: r3 (X); r3 (Y); w3 (Y);

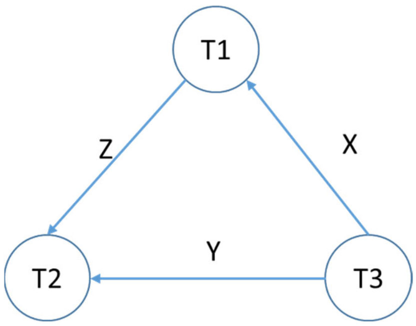
S1: r1(X); r2(Z); r1 (Z); r3(X); r3(Y); w1(X); c1; w3(Y); r2(Y); c3; w2(Z); w2(Y); c2;

S2: r1(X); r2 (Z); r3 (X); r1(Z); r2 (Y); r3 (Y); w1 (X); c1; w2 (Z); w3 (Y); w2 (Y); c3; c2;

1) Draw the serialization graphs for S1, S2 and state whether each schedule is serializable or not. If a schedule is serializable, write down all equivalent serial schedule(s). [20 marks]

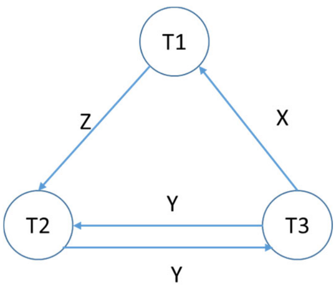
Solution:

S1:




It is serializable because the graph doesn't contain a cycle. The only equivalent serial schedule is T3 -> T1 -> T2.

S2:




The schedule isn't serializable because there is a cycle.

2) Please determine the strictest recoverability condition (non-recoverable, recoverable, cascadeless or strict) that each schedule satisfies.[10 marks]

Solution: S1 is recoverable since T2 reads Y which is written by T3, and T3 commits before T2 commits. But it is non-cascadeless, since when T2 reads Y, T3 has not committed yet. [5 marks] S1 is recoverable since T2 reads Y which is written by T3, and T3 commits after T2 commits. But it is non-cascadeless, since when T2 reads, T3 has not committed yet.

S2 is cascadeless, because every transaction reads only the items that are written by committed transactions. But it is not strict since T3 writes Y before T2 but when T2 writes Y T3 has not committed yet.