



North South University

Department of Electrical & Computer Engineering (ECE)

Final Assignment

CSE311 (Database Systems) Fall 2021

15 Points

Submission Portal: Canvas

Name

Mohammad Saifur Islam

Student ID

1922071042

This assignment is worth 15% of your marks in the subject.

This paper has 21 pages.

Name : Md. Saifur Islam Student ID : 1922071042

Instructions:

1. Write Name, Student ID on top of each page.
2. You need to print this assessment file.
3. Write answer for each question in the designated space.
4. Show your work. Writing just your final answer is not enough.
5. Pages without name and Student ID will not be marked.
6. This is an individual work. Any unauthorized help will result in cancellation of the script. Any examinee found adopting unfair means will be expelled from the trimester/program as per NSU disciplinary rules.
7. You should answer all questions. Write in your own words. Do not copy & paste from web.
8. Answers must be written in blue or black ink.
9. Textual answers can be in point form.
10. Submit the answer script in PDF format before due date. No late submission will be allowed. Any late submissions will be designated as incomplete.
11. Proven case of direct code copies from other students, or online sources will result in severe point deduction.

Write and sign the Honor Code pledge.

"On my honor I affirm that I have neither given nor received inappropriate aid in the completion of this exercise."

On my honor I affirm that I have neither given nor received inappropriate aid in the completion of this exercise.

X Mohammad Saifur Islam

Student

Name : Md. Saiful Islam Student ID : 1922071042

1. We are modelling a database for storing students and their marks. Students have an id number and name, and receive marks for a series of assignments, each of which has an id number and a title. A student does a given assignment only once – we record the date that the student submitted the assignment.

Our modeller has arrived at the following relation (primary key is underlined).

STUDENTMARKS (studentid, givenName, surname (assignmentId, assignmentTitle, mark, dateSubmitted))

But this relation is not in third normal form.

First, explain why it is not.

Second, convert the relation to 3rd normal form.

Mark your primary keys with a solid underline, and your foreign keys with a dotted underline.

STUDENMARKS (studentid, givenName, surname
(assignmentId, assignmentTitle,
mark, dateSubmitted))

First we write the functional dependency for this case.

$\text{assignmentID} \rightarrow \text{studentid, assignmentTitle, mark, givenName, surname}$

$\text{studentid} \rightarrow \text{givenName, surname}$

In the above functional dependency, there occurs a transitive dependency, where

$\text{assignmentID} \rightarrow \text{studentID}$ and
 $\text{studentid} \rightarrow \text{givenName, surname}$

Name: Md. Sufiul Islam Student ID: 1922071012

In order to achieve 3NF, eliminate Transitive Dependency. This will end with the following two tables:

Student (StudentID, GivenName, Surname)

Assignment (AssignmentID, AssignmentTitle, marks, date, Submitted, StudentID (ref. key))

Name : Md. Sufiul Islam Student ID : 1922074042

2. A) You are designing a database for a new social media startup. Give three (3) reasons why you might choose a distributed database. (5 Mark)

Three reasons of choosing distributed database are:

- (i), Better Reliability: If a failure happens in a particular area of a distributed database then the entire social media site will not crash; instead only a portion of the website will be affected.
- (ii), Easier Expansion: Expansion of data in distributed database by adding more data or adding more processor is easier.
- (iii), Improved Performance: Distributed database is faster as we can fetch several queries from various portion of the DB at the same time.

B) Describe the differences between replication and partitioning. (5 Marks)

Replication	Partitioning
(i), copy data among several nodes on the internet.	(i), split database into smaller parts across the internet.
(ii), multiple nodes can have same set of data.	(ii), multiple nodes cannot have same set of data.
(iii), simple in implementation.	(iii), complex in implementation.
(iv), It's more reliable because even if one node fails, data can still be retrieved from other nodes.	(iv), It's less reliable since if a node goes down, the data can't be retrieved at all.

Name: Md. Sufiul Islam Student ID: 1922071042

C) List four (4) possible disadvantages of using a distributed database. (5 Marks)

Four possible disadvantages of using distributed database are:

- (i) Complex nature: The nature of distributed database is more complicated than that of a centralised database. Distributed database systems necessitate complex software. It also assures that no data is replicated, adding even more complexity to the system.
- (ii) Increased cost: Various expenses such as maintenance, procurement, hardware, network costs, labour costs and so on, add up to a higher total cost than normal database systems.
- (iii) Security concern: In a distributed database along with maintaining no data redundancy, the security of data and a network is a prime concern. A network can be easily attacked for data theft and misuse.
- (iv) Integrity control: maintaining data consistency is critical in a vast distributed database. Changes made at one site must be mirrored across all sites. To ensure data integrity, transmission and processing costs are high in distributed database system.

Name : Md. Saiful Islam Student ID : 1929071042

3. An engineering consultancy firm supplies temporary specialized staff to bigger companies in the country to work on their project for certain amount of time. The table below lists the time spent by each of the company's employees at other companies to carry out projects. The National Insurance Number (NIN) is unique for every member of staff.

NIN	Contract No	Hours	Employee Name	Company ID	Company Location
616681B	SC1025	72	P. White	SC115	Belfast
674315A	SC1025	48	R. Press	SC115	Belfast
323113B	SC1026	24	P. Smith	SC23	Bangor
616681B	SC1026	24	P. White	SC23	Bangor

- i) Explain in which normal form this table is
- ii) Find the Primary Key for this relation and explain your choice.
- iii) Find the Functional Dependencies on the PK.
- iv) Normalise the table to 2NF.
- v) Find the transitive dependencies on the 2NF tables
- vi) Normalise the tables to 3NF and show the PK and FK in all the relations.

(i) Table is in 1NF because there is no multivalued attribute.

(ii) NIN + contract No can be used as primary key for this table. Because NIN alone is not unique in all the rows nor is contract no alone.

Name: Md. Sajidul Islam Student ID: 1922071042

(iii) Functional Dependencies:

$NIN, contractNo \rightarrow hours, EmployeeName, companyID$

$companyID \rightarrow companyLocation$

$NIN \rightarrow EmployeeName$

$contractNo \rightarrow companyID$

(iv) Dividing and Underlining Primary Key (2NF)

staff (NIN , EmployeeName)

contract (contractNo, companyID)

work (NIN , contractNo, hours)

(v) Table contract contains transitive dependency

$companyID \rightarrow companyLocation$

(vi) staff (NIN , EmployeeName)

contract (contractNo, companyID. (FK))

company (companyID, companyLocation)

work (NIN , contractNo, hours)

Name: Md. Saiful Islam Student ID: 1922071049

4. Explain why, when implementing a replicated NoSQL database, we must trade-off between availability and consistency.

The CAP theorem states that a distributed database system has to make a trade-off between consistency and availability when a partition occurs.

We can not guarantee three of them.

NoSQL database are classified based on two characteristics:

(a) CP: It delivers consistency and when a partition occurs in any nodes, the system has to shutdown until the partition has been resolved.

(b) AP: It delivers when availability when a partition occurs all nodes remain available but some nodes may return older version of data.

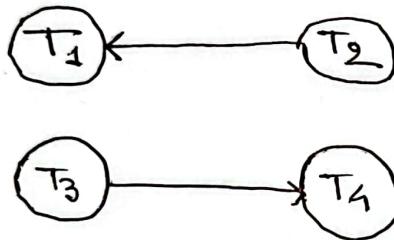
So, we see that if we want to offer availability we can not guarantee consistency and vice versa. That is why we must trade-off between availability and consistency.

Name: Md. Saiful Islam Student ID: 1922071042

5. Draw the Precedence Graph for the schedule and indicate if it is conflict serializable (if yes write down a conflict equivalent to the schedule)?

	1	2	3	4	5	6	7	8	9	10
T_1	R(A)						W(A)	R(C)		
T_2		R(A)							R(C)	
T_3				R(C)	R(B)					W(B)
T_4			R(C)			W(C)				

Precedence Graph:



No cycle in precedence graph.

So, it is conflict serializable

conflict equivalent schedule

$\rightarrow \underline{T_2} \underline{T_3} \underline{T_1} \underline{T_4}$

Name: Md. Saiful Islam Student ID: 1992074042

6. Explain, using a timeline, how locking can prevent the "Lost Update" problem.

One recommended method for preventing lost updates is to use optimistic concurrency control to perform what is called optimistic on the data. Optimistic concurrency control type uses four phases in order to help to ensure that data doesn't lost:

(i) Begin: A time stamp is recorded to pinpoint the beginning of the transaction.

(ii) Modify: Read values and make write tentatively.

(iii) Validate: Make a check to ensure that other transaction have not modified any data that is used by the current transaction.

(iv) Commit: If there are no conflicts, then transaction can be committed. Otherwise the transaction can be aborted to prevent a lost update from occurring.

Name: Md. Saiful Islam Student ID: 1922071042

7. Describe the four properties of transactions referred to by the acronym "ACID".

The acronym "ACID" refers to the four key properties of a transaction:

- (i) Atomicity: All the changes to data are performed as if they are a single operation. That is all the changes are performed or none of them are.
- (ii) consistency: Data is in a consistent state when a transaction starts and when it ends.
- (iii) Isolation: The transaction should be isolated, that is, the execution of a transaction should not be related with any other transactions executing concurrently.
- (iv) Durability: changes to data endure a transaction completes successfully and are not undone, even if the system fails.

⇒ Example: In an application that transfers fund from one account to another,

1. atomicity → ensure debit is made successfully from one account.
2. consistency → total value both account same at start and end.
3. isolation → ensure another transaction sees the transferred funds in one/other but not in both, nor in neither.

4. Durability → ensure changes made to each account will not be reversed.

Name : Md. Saiful Islam Student ID : 1929071042

8. Describe two (2) problems with relational databases that NoSQL databases aim to address.

Two problems are :

1. The first problem with relational databases is that they require a schema to be created before data can be entered. But in NoSQL, there are predefined schemas and they do not require that should be defined for storing and managing the data.
2. In relational database, the data can only be stored in a table as a rows and columns. whereas in NoSQL, the data can be stored in any format such as documents, key-values stores, graphs etc.

Name : Md. Saiful Islam Student ID : 1922071092

9. Excel Airlines operates flights between major Australian cities. It currently keeps its schedule in a spreadsheet: a sample of the spreadsheet is shown below. In this sample, flight 123 flies daily at 11am from Melbourne to Sydney, while flight 44 flies on weekends at 2pm from Sydney to Melbourne.

FlightNumber	timeTakeoff	fromAirportCode	fromAirportName	toAirportCode	toAirportName	dateTakeoff	pilotId	pilotName
123 11am		MEL	Melbourne	SYD	Sydney	Fri 3rd Nov	1	Dave
123 11am		MEL	Melbourne	SYD	Sydney	Sat 4th Nov	2	Marie
123 11am		MEL	Melbourne	SYD	Sydney	Sun 5th Nov	3	Ibrahim
44 2pm		SYD	Sydney	MEL	Melbourne	Sat 4th Nov	1	Dave
44 2pm		SYD	Sydney	MEL	Melbourne	Sun 5th Nov	3	Ibrahim

The company would like to start storing their data in a relational database. Their initial data model, based on the spreadsheet, is expressed in relation notation as:

Flight(flightNumber, timeTakeoff, fromAirportCode, fromAirportName, toAirportCode, toAirportName, (dateTakeoff, pilotId, pilotName))

However, this data model is not in 3rd normal form. Your job is to convert it to 3rd normal form. You don't need to show intermediate normal forms – just the 3rd normal form you end up with. Mark your primary keys with a solid underline, and your foreign keys with a dotted underline. (Any attributes that are both primary and foreign keys should get both underlines.)

3NF of given relation:

- Primary keys are given solid underline
- Foreign keys are given dotted underline
- Both primary and foreign key should get both underlines.

pilot(pilotId, pilotName)

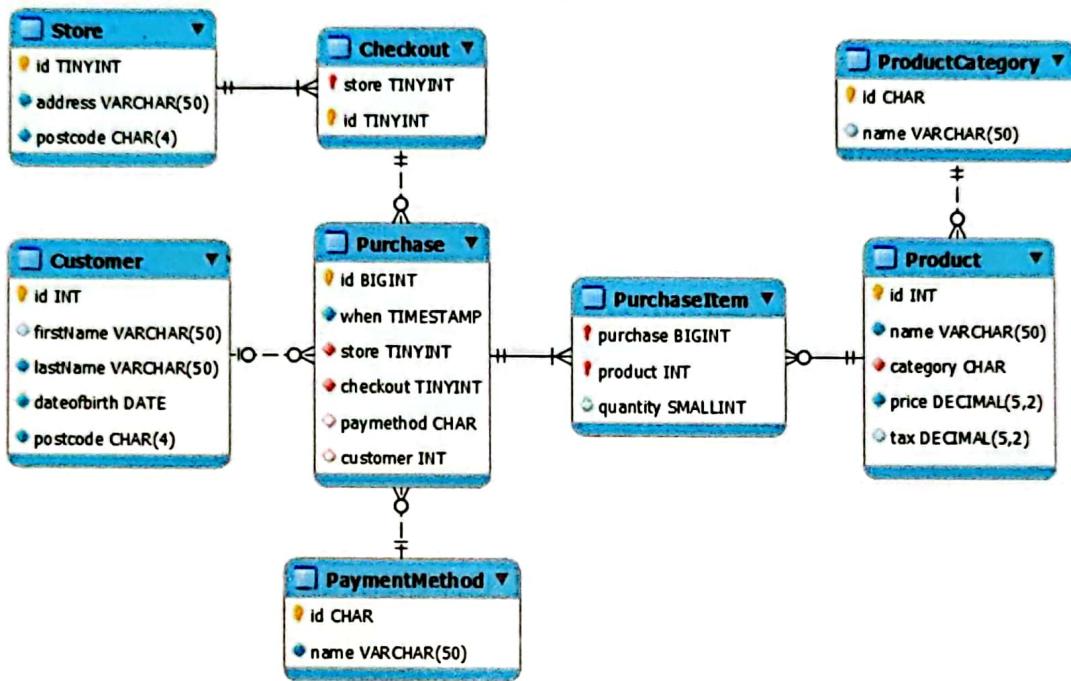
airport(airportCode, airportName)

flight(flightNumber, timeTakeoff)

flight-info(flightNumber, fromAirportCode, toAirportCode, dateTakeoff, pilotId)

Name : Md. Safful Islam Student ID : 1992071042

10. Consider the following database for a chain of 5 supermarkets in Melbourne. Customers take a collection of products to a checkout for scanning: this is one "Purchase". Each item within the purchase is a "PurchaseItem". Some customers identify themselves at the checkout by scanning a loyalty card. The checkouts are numbered within each store: so a given store n has checkouts $n-1, n-2$ etc. Each product is classified within a particular category. There are several payment methods. The following ER diagram describes the database schema which has been implemented.



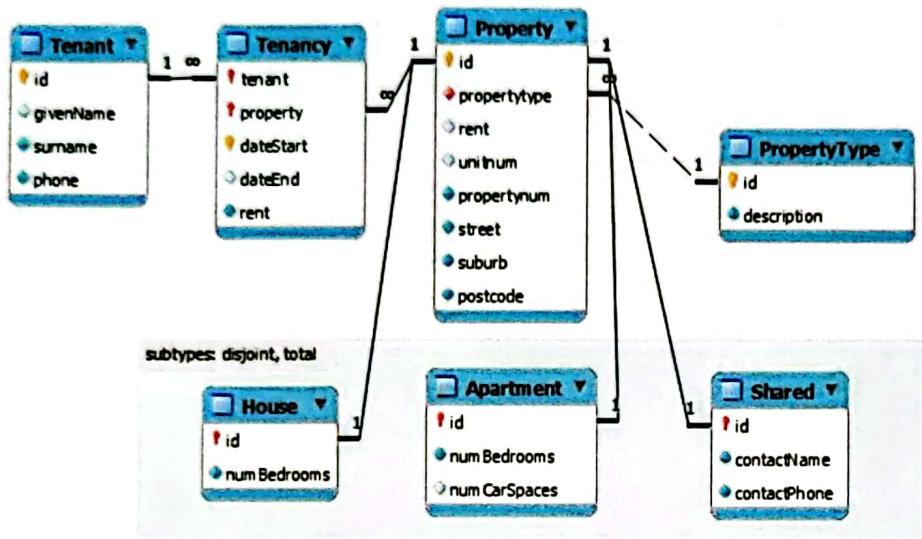
it has been found that the most common queries being run against the Product table are the following:
 select price from Product where id = @input;
 select price, tax from Product where name = @input;
 Based on this evidence, which columns in the Product table should we index?

We must index id and name on this occasions.

Name : Md. Sufiul Islam Student ID : 1922071049

11. Consider the following data model and *sample data* (not all data is shown) for a student real-estate system.

At any given time, an individual property is empty or occupied by one tenancy. Note that the rent charged for a particular tenancy, especially one in the past, may be different to the current advertised rent on that property.



id	givenName	surname	phone
1	Britteny	Abston	0441532370
2	Margy	Alter	0810478134
3	Freida	Amaral	072681647
4	Kera	Anderson	0266223748
5	Barbra	Basham	0353849916
6	Lon	Batchelor	0924374652
7	Heide	Belew	0987544272
8	Susan	Bergen	083194347
9	Ouda	Binion	0599985340
10	Wai	Bisceglia	0606937760

Tenant	Property	dateStart	dateEnd	rent
92	5	2014-08-09	NULL	429
160	13	2014-04-23	2016-04-22	362
153	39	2014-01-11	2015-01-11	339
48	42	2014-05-25	2015-05-25	291
93	48	2014-06-14	NULL	314
79	49	2014-02-18	2015-02-18	517
36	51	2014-01-05	2015-01-05	395
132	51	2014-08-11	NULL	425
6	82	2014-06-06	2015-06-06	384
86	123	2014-04-07	2015-04-07	458
109	133	2014-03-30	2015-03-30	534

id	PropertyType	rent	unitnum	propertynum	street	suburb	postcode
1	H	394	NULL	39	Napier lane	Fitzroy	3065
2	H	404	NULL	97	Charles place	Fitzroy	3065
3	H	540	NULL	68	Harrison place	Fitzroy	3065
4	H	590	NULL	48	Royal lane	Fitzroy	3065
5	H	429	NULL	38	Johnston street	Fitzroy	3065
6	A	374	4	45	Rodney place	Carlton	3053
7	H	586	NULL	12	Bouverie street	Carlton	3053
8	S	107	NULL	26	Kerr street	Fitzroy	3065
9	A	378	4	93	Hightett place	Fitzroy	3065
10	S	171	NULL	90	Charles place	Fitzroy	3065

id	description
A	Apartment
H	House
S	Shared

Name : Md. Saifur Islam Student ID : 1922071042

id	numBedrooms
1	2
2	3
3	4
4	4
5	3
7	4
11	3
13	2
14	3
17	3

id	numBedrooms	numCarSpaces
6	1	2
9	1	1
12	1	0
15	1	2
21	1	1
24	2	1
30	2	2
33	1	0
36	1	0
39	1	1

id	contactName	contactPhone
8	Finley	8697660699
10	Jovanny	1597228831
16	Deborah	2319846495
20	Dean	8566163983
28	Stanley	9176265253
52	Krish	4854635066
64	Haiden	6938385335
76	Alvaro	7457141435
88	Dylan	6118526965
106	Destiny	7316239255

Questions A-D require you to write SQL statements. Your code should be formatted so it is easy to understand. Ensure user friendly output by displaying and sorting on appropriate fields and naming fields appropriately.

A) Which tenants have paid us the most rent?

List the top 5 payers, and how much rent each has paid in total.

```
WITH rank_rent AS (
SELECT te.tenant as tenant, tt.givenName as Name, sum(te.rent)
as total_rent, RANK() OVER (PARTITION BY te.tenant,
tt.givenName)
ORDER BY sum(te.rent) DESC) rent_value
FROM tenancy te INNER JOIN tenant tt
ON te.tenant = tt.id)
SELECT * FROM rank_rent WHERE rent_value <= 5;
```

B) What is the longest given name among our tenants?

If several names tie for first place, list them all.

```
SELECT givenName
FROM Tenant
WHERE length(givenName) IN (SELECT MAX(length(givenName))
FROM Tenant);
```

Name : Md. Saiful Islam Student ID : 1922071042

C) List the addresses of the houses and apartments which only have one bedroom.

```
SELECT p.street as street, p.suburb as suburb, p.postcode as postcode
FROM house h JOIN property p
ON (h.id = p.id)
WHERE h.numBedrooms = 1
UNION
SELECT p.street as street, p.suburb as suburb, p.postcode as postcode
FROM apartment a JOIN property p
ON (a.id = p.id)
WHERE a.numBedrooms = 1;
```

D) List the names of tenants who have rented at least two different types of properties.

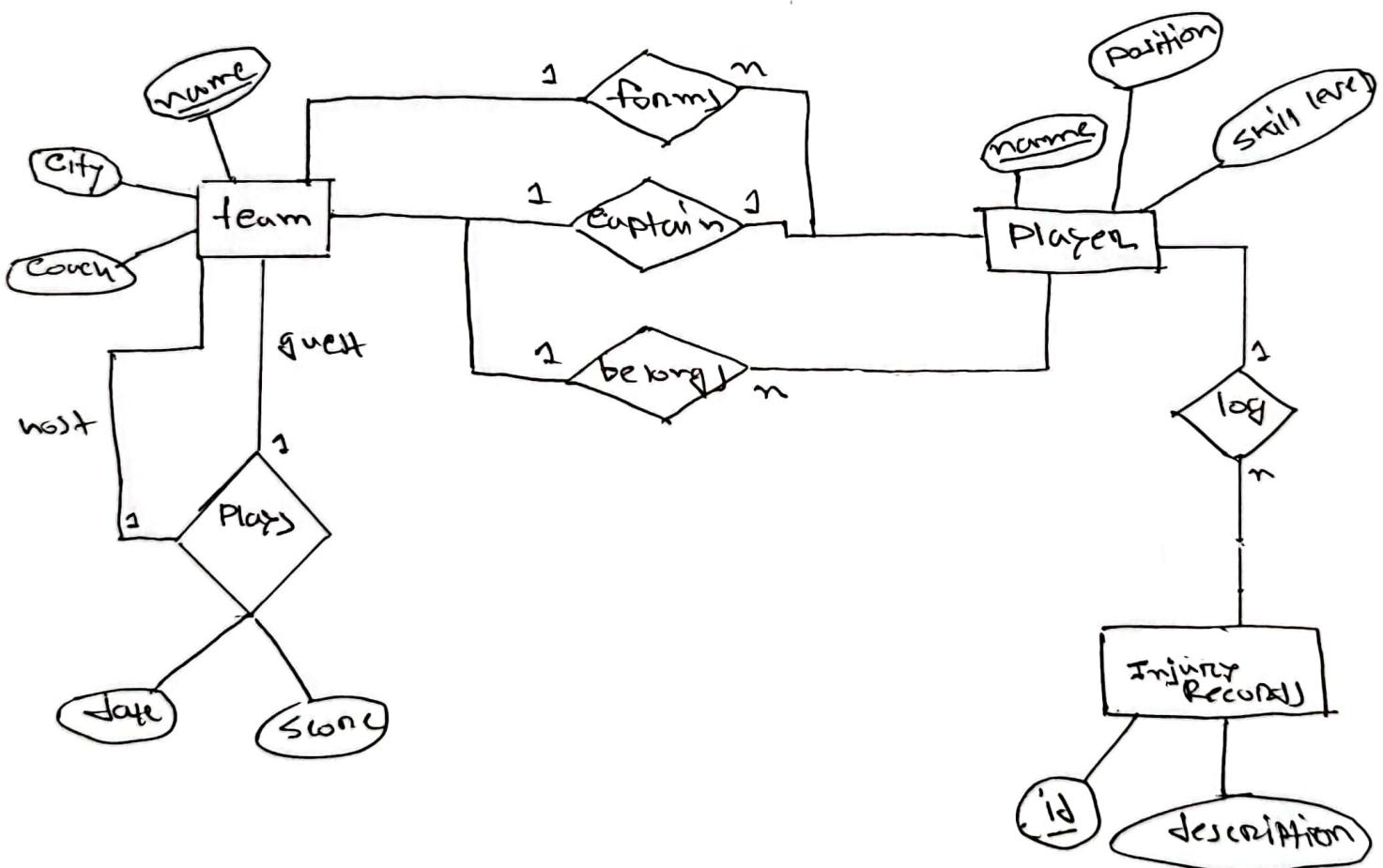
```
SELECT t.id as id, t.givenName as name
FROM tenant t JOIN tenancy te
ON (t.id = te.tenant)
JOIN property p
ON (te.property = p.id)
GROUP BY t.id, t.givenName
HAVING COUNT(DISTINCT p.propertyType) >= 2;
```

Name : Md. Sufiyan Islam Student ID : 1922071042

12. Suppose you are given the following requirements for a simple database for the National Hockey League (NHL):

- The NHL 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 a name, a position (such as left wing or goalie), a skill level, and a set of injury records,
- A team captain is also a player,
- A game is played between two teams (referred to as host_team and guest_team) and has a date (such as May 11th, 1999) and a score (such as 4 to 2).

Construct a clean and concise ER diagram for the NHL database.



Name: Md. Sufiul Islam Student ID: 1922071042

13. Discuss differences among functions, procedures, and triggers in SQL. Explain briefly with different scenarios/examples where we might use them.

Difference between procedures, Functions and Triggers:

1. In Execution:

~~~~~

(i) Procedure: We can execute the stored procedure

when required.

(ii) Function: We can call a function whenever required. Function can't be executed because a function is no pre-compiled form.

(iii) Trigger: Triggers may be set to run automatically when a table is updated, deleted or inserted.

### 2. Calling:

~~~~~

(i) Procedure: Procedures cannot be invoked from a function, may be called from select statement but procedure cannot. Procedure, on the other hand may be called from trigger.

(ii) Function: Function can be called from procedure or trigger.

(iii) Trigger: can't be called from procedure or trigger.

3. Parameter:

(i) Procedure: can accept any type parameter.
Also accept out parameter.

(ii) Function: can accept any parameter but function can't accept out parameter.

(iii) Trigger: can't pass parameter.

4. Return:

(i) Procedure: may or may not return any value.

(ii) Function: must return any value

(iii) Trigger: never return value.

Example: Assume the overall expense computation is dependent on a set

of criteria; instead of writing the logic numerous time, you may invoke a function. A function can only return one value, which is usually all you need.

You can utilize a procedure with arguments if you require more than one value returned.

While triggers provide a very different purpose. The trigger could be helpful if you want to run a series of statements when a specific condition is fulfilled triggering event.

Name: Md. Sufiyan Islam Student ID: 1922074042

This page is left blank intentionally.