EECS 484 Database Management Systems
Homework 1
Due on Friday, September 13th @ 11:45 PM Eastern time

Please read the following instructions before starting the homework:

There are two parts to this homework.

The first part consists of short response questions.

The second part requires you to draw out an ER diagram. Please use a high quality camera or a scanner and scan your answers into one PDF file or use a computer tool (e.g., Lucidchart) to draw the diagrams. Be sure to check your PDF submission to ensure that your ER diagrams are readable.

This homework must be completed individually and can be submitted on Gradescope. Use entry code **XG8VVB** to self-enroll if you don't have access to the Gradescope course page.

No late days for homework! *If you miss the due date, you get 0 points. If your PDF gets modified after the due date, you get 0 points. No exceptions on this.*

Honor Code

By submitting this homework, you are agreeing to abide by the Honor Code:
I have neither given nor received unauthorized aid on this assignment, nor have I concealed any violations of the Honor Code.

PART 1 – Short Responses

Question 1 (2 points)

Determine if the following statements describe a conceptual/logical, physical, or external representation/schema of data.

a) Student records contain personal information such as their name, uniqname, email, social security number, and home address in Wolverine Access. However, professors cannot see a student's social security number and home address when viewing student records.

   a. conceptual/logical
   b. physical
   c. external

b) Data entries are stored across binary files on a private AWS server.

   a. conceptual/logical
   b. physical
   c. external

Question 2 (8 points)

Use the following instance (i.e. snapshot/subset) of a cars relation to answer the questions below.

cid	*brand*	*model*	*year*	*miles_per_gallon*
35692	Ford	Ford Escape	2018	33
48577	Acura	Acura RDX	2017	28
8721	Jeep	Jeep Wrangler	2017	23
5763	Jeep	Jeep Grand Cherokee	2016	21
35871	Honda	Honda CR-V	2022	34
23190	Toyota	Toyota Corolla	2024	38

Just using the information in the given *instance* of a cars relation, identify the following attributes (set) as *might be a key* or *definitely not a key* for the **entire** cars relation, and provide a brief explanation.

1. miles_per_gallon definitely not a key, because MPG may not be unique
2. cid might be a key. because cid is set unique for each car
3. brand definitely not a key, because there are two Jeep
4. {brand, model} might be a key, because the combination of brand and model is unique.

Answer the following questions, just based on the information in the given instance:

5) Which single attribute is **definitely** a candidate key? Why? If there are no candidate keys, explain why.

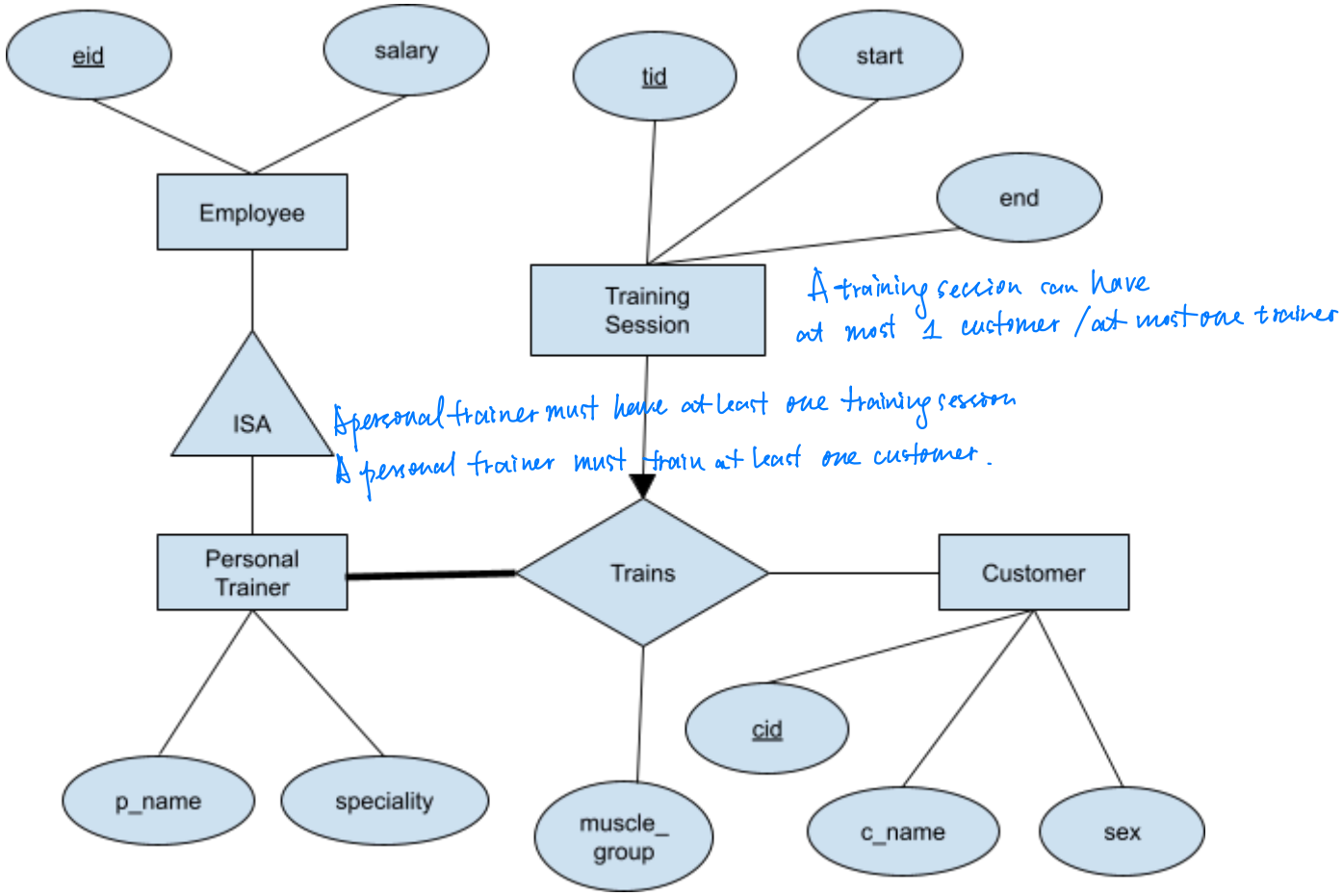
cid. Because it uniquely identifies each row in the relation.

6) What is a reasonable logical (conceptual) schema for the above relation?

Car (cid: integer, brand: string, model: string, year: integer, MPG: integer)

Question 3 (14 points)

The following ER diagram shows the relationships between a personal trainer and customers at a gym.



Determine whether each of the following is true or false, given the constraints reflected by the above ER diagram. No justification is required. (2 points each)

a) There can be a personal trainer that does not train any customers.
   True / False

b) Every customer must have a personal trainer.
   True / False

Questions continue onto the next page.

c) The Personal Trainer entity does not have a primary key.
True / False

d) A personal trainer, Sam (eid=24), can train two customers Ellie (cid=18) and Riya (cid=7) in two different training sessions (tid=3 and tid=4).
True / False

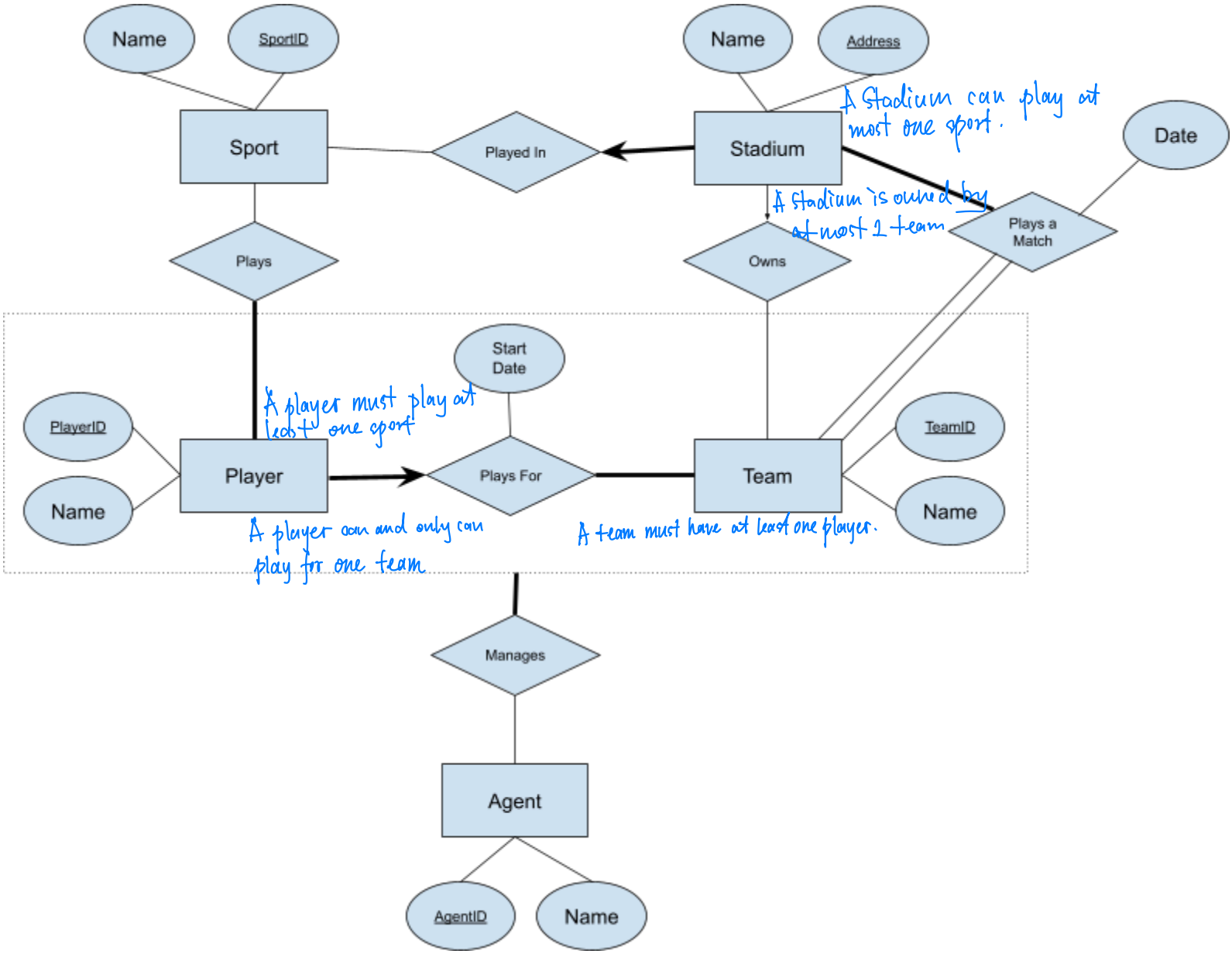
e) There can be training sessions that do not have any customers being trained.
True / False

f) A customer, Max (cid=9), can train two different muscle groups, legs and arms, in the same training session (tid=1) with the same personal trainer Gabe (eid=27).
True / False

g) Both personal trainers Braden (eid=4) and Zach (eid=6) can train customer Lili (cid=12) in the same training session (tid=10).
True / False

Question 4 (20 points)

Consider the following ER diagram that represents one way a multiple-sport database could be designed and answer the questions on the next page.



Answer the following true/false questions referring to the previous ER diagram. Justification is not required. (2 points each)

a) There can be stadiums in which multiple sports are played.
True / False

b) A player, Nick (pid = 14), can play for multiple teams.
True / False

c) A team can consist of players that play different sports.
True / False

d) Every agent must manage a relationship between a player and team.
True / False
at least 1.

e) A player, Rory (pid = 4) can play for the same team (tid = 2) with two different start dates.
True / False

f) A relationship between a player, David (pid = 9) and their team, the Titans (tid = 6), can be managed by multiple agents.
True / False

g) There can exist sports that contain no players that play it.
True / False

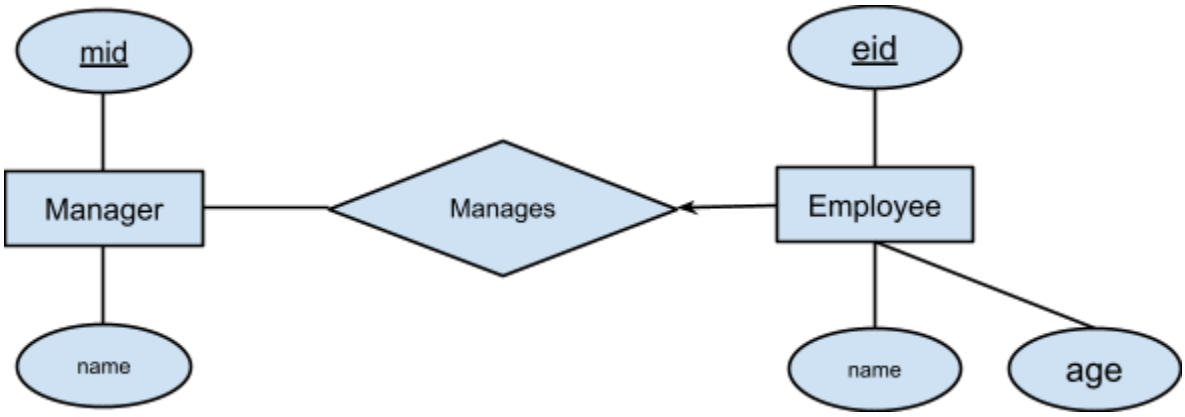
h) A match between two teams requires all the players on those teams to play the same sport.
True / False

i) A team such as the Cougars (tid = 17) can maintain ownership of multiple stadiums.
True / False

j) Two teams can play a match against each other as many times as they would like, as long as each match is in a different stadium.
True / False

Question 5 (14 points)

Consider the following ER diagram and answer the following questions about keys.



The following questions are about table representation of the above diagram.

a) The fewest number of tables that can be used to represent this ER diagram *without redundancy* is 3 (meaning that we need at least 3 tables to represent this diagram) (2 points)
True / False

b) In the least redundant design, 'mid' is a foreign key in the Employee table (2 points)
True / False

c) A candidate key for the 'manages' relationship is {eid} (2 points)
True / False

d) A candidate key for the 'manages' relationship is {mid} (2 points)
True / False

e) In the least redundant design, 'eid' is a foreign key in the Manager table (2 points)
True / False

Part f is on the next page.

Create a table for **Employee** by SQL statement including all the necessary integrity constraints shown by the above ER diagram (4 points)

```
f) CREATE TABLE Employee (
        eid   INTEGER PRIMARY KEY
        name  CHAR(30)
        age   INTEGER
);
```

PART 2 – ER Diagrams

Question 6 (36 points)

As a basketball lover and an EECS484 enthusiast, you want to design a database to model national basketball teams. This includes the teams themselves, the games they play, the players and the referees.

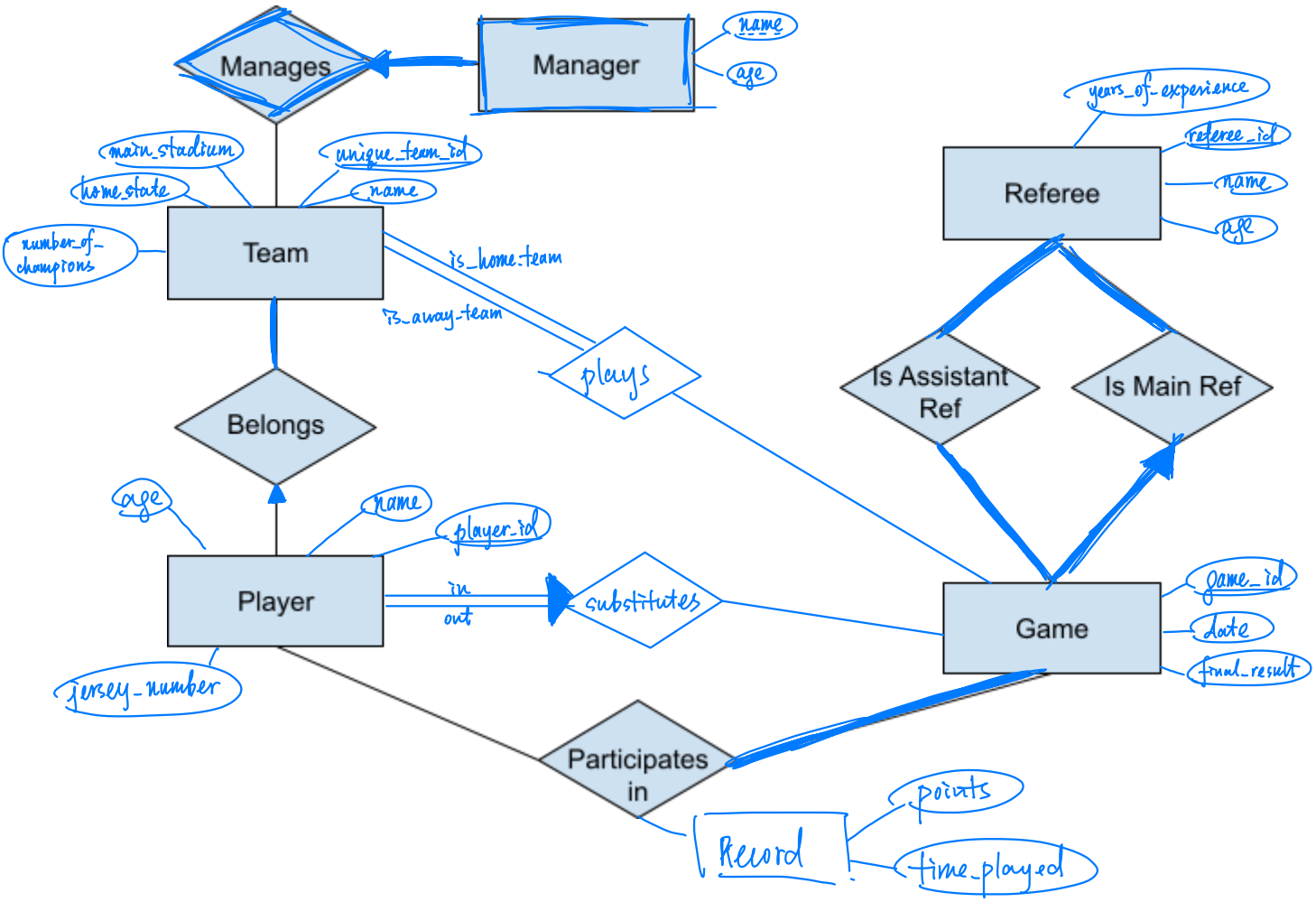
In the design, you want to capture the following:

- For each basketball team, you want to store their unique team ID, name, main stadium, their home state, and the number of championships they've won.
- Each player on each team will have a unique ID, name, age, and jersey number.
    - Each team has at least one player and each player can belong to only one team. *It is possible that a player does not belong to any team.*
- Each manager on each team will have a name and age. Each manager manages only one team and each manager's name identifies them uniquely from among the team's administrators. If a team is deleted, you need not keep track of its manager any longer.
- You want to store data on all basketball games:
    - Each game must have exactly one away team and one home team.
    - Each game has a game ID, date, and a final result. game ID is unique to each different game.
    - The players that participated in a game (each game must have more than one player) must be stored, as well as:
        - The number of points the player scored, and their total time playing in that game
    - Players can substitute for other players during each game, which should be recorded as well:
        - In one game, one player may substitute in or out.
        - In one game, the same substitution can occur only once.
        - Any number of substitutions can be made per game.
        - A player cannot substitute for themselves.
- You also want to capture information about the referees for each game to analyze their performance and potential biases, so you'd also like to capture the following:
    - Each game has multiple referees. Each referee should have an ID, name, age, and years of experience. The IDs for different referees are unique.
    - There is exactly one main referee and one or more assistant referees per game.
    - Each referee works as a main referee in at least one game, and as an assistant referee in at least one game.
    - A referee in one game works as either main referee or assistant referee but not both.

1) Design an ER diagram that best reflects the constraints described above. You can start with the ER diagram below. Always follow the constraints described in the questions, even if you don't think they make sense in the real world. Try to reflect as many constraints as possible. Some constraints may not be possible to express in an ER diagram. Do not make any arbitrary assumptions or modify the constraints given in the problem statement.

Make sure that all your entities, attributes and constraints are drawn (again, state assumptions clearly if they are not stated in the question). Please do not modify any given entities and attributes. (32 points)

HINT: To complete this question, you may want to add more entities, relations and attributes. You may want to add constraints in the given diagram as well.



2) It is possible that some constraints mentioned above cannot be enforced by the ER diagram features we learned in lecture. State these constraints. (4 points)

1. Each referee works as a main referee in at least one game, and as an assistant in at least one game.

2. Each game must have more than one player (currently it is at least 1 player).