University of Manitoba

Designing a Database

Data Discovery and Database Design

COMP 3380: Databases Concepts and Usage

Authors:
Garik Avagyan - 7893763
Ivan Balkashynov - 7900097
Tim Sarvis - 7866126
Uday Shekhawat - 7923995

Instructor:	*Anifat Olawoyin*
Department:	Computer Science
Date:	March 1, 2023

Contents

1 Brief Summary of Dataset

The datasets used in this project are a combination of publicly available datasets through services like Kaggle and the manual addition of data to fill out some gaps in those datasets. There are three main datasets which will be drawn from: players information, team information, and awards information.

The players information dataset will contain information about all NBA players and their stats in the year 2021. Specifics like draft records, season statistics, colleges players played for, and personal details about each NBA player are recorded in this dataset. This dataset contains 605 players(rows) with 21 columns.

The team information dataset will contain information about all NBA teams and their personnel, season statistics, and owners in the year 2021. Specifics like where the team is located, who coaches the team, who owns the team, team records during the season, and what league the team is in are recorded in this dataset. There are 30 rows(teams) with 20 columns.

The awards information dataset will be a small dataset that contains the information about the various league awards and who won them in the year 2021. This includes specific awards such as MVP, rookie of the year, and more. There are 7 rows(awards) with 2 columns.

From part B, this dataset will break down to 12 tables with around 2000 rows.

2 ER Model

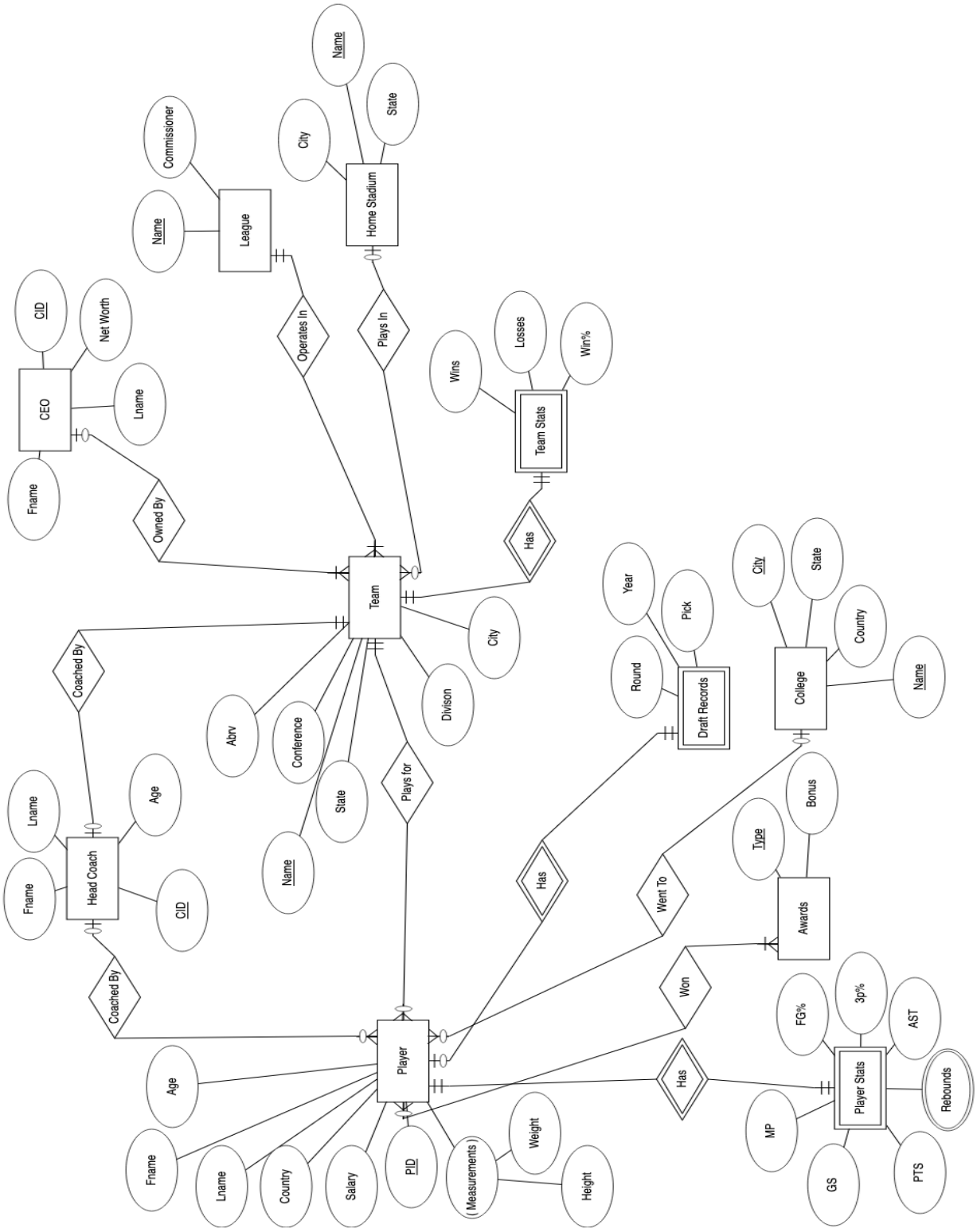

Figure 1: ER Model

Explanation/Assumptions of Various Components

- Rebounds can be a multi-valued attribute since they can include offensive and defensive rebounds.
- For US Colleges, multiple colleges with the same name cannot exist in the same city so Name,City is the primary key.
- For Team, no NBA team can have a duplicate name so name is a suitable primary key. Similar ideas are used for the league and home stadium entities.
- Draft records, Stats, and Team Stats have no partial keys

3 ER Model Participation and Cardinality

The following relationships are contained within the EER model:

- Player **Plays_For** Team
  - Cardinality: A player can only be under contract for one team and each team can have 15 players under contract so many players
  - Participation: A player can exist without playing for a team by being a free agent and a team cannot exist without players playing for the team.
  - Players: Many with optional
  - Team: One with total
- Player **Coached_By** Head Coach
  - Cardinality: As per NBA structure, there is only one head coach per team. This means that there is one coach for the many players on an NBA team (15).
  - Participation: A player can exist without a coach by being a free agent and similarly a coach can exist without a player by being a free agent coach.
  - Players: Many with optional
  - Team: One with optional
- Player **Has** Stats
  - Cardinality: Each player has a group of specific stats tied to them and recorded for them throughout the season. Therefore one player has one set of stats that's unique to them for that season.
  - Participation: Each player must have their stats officially recorded and stats cannot exist without a player. Therefore both have total participation.
  - Player: One with total
  - Stats: One with total
- Player **Won** Awards
  - Cardinality: One player has the possibility to win multiple awards in a given season or each award can go to multiple different players. Therefore many players can win many awards.
  - Participation: A player can exist without winning an award, but an award needs to be given to a player as per NBA rules.
  - Player: Many with Optional
  - Award: Many with Total

- Player **Has** Draft Records
  - Cardinality: Each player has a group of specific draft records tied to them. Therefore one player has one set of draft records that is unique to them
  - Participation: Each player does not need draft stats since they can go undrafted, but draft stats need a player to exist.
  - Player: One with optional
  - Draft Stats: One with total
- Player **Went_To** College
  - Cardinality: Many players can come from the same college. It is assumed that each player can only go to one given college(the one their drafted from)
  - Participation: Each player does not need to go to college since they can get drafted from high school. A college can exist without NBA players.
  - Player: Many with optional
  - College: One with optional
- Team **Coached_By** Coach
  - Cardinality: As per NBA structure, there is only one head coach per team. Therefore this is one coach per NBA team.
  - Participation: A coach can exist without a team as a free agent, but a team cannot exist without a head coach as per NBA rules.
  - Team: One with total
  - Coach: One with optional
- Team **Owned_By** Owner
  - Cardinality: Each team has only one CEO that bought the team. A CEO can theoretically buy multiple teams and become CEO of each team.
  - Participation: Each team needs a CEO to make business decisions. A CEO can exist without owning an NBA team.
  - Team: Many with total
  - Owner: One with optional
- Team **Operates_In** League
  - Cardinality: Each team can only be legally part of one league. Each league can house many teams(the NBA houses 30).
  - Participation: A team cannot exist without being in a professional league. A league cannot exist without teams playing in the league.
  - Team: Many with total
  - League: One with total

- Team **Plays_In** Home Stadium
  - Cardinality: Each team has a home stadium that they must play their home games in. Each stadium can have multiple teams as their home team(Lakers and Clippers).
  - Participation: A team can exist without a home stadium to play in(such as in COVID when they played in a bubble) and a stadium can exist without an NBA team.
  - Team: Many with optional
  - Stadium: One with optional
- Team **Has** Stats
  - Cardinality: Each team has a group of specific stats tied to them. Therefore one team has one set of stats that is unique to them.
  - Participation: Each team must have their stats officially recorded and stats cannot exist without a team. Therefore both have total participation.
  - Team: One with total
  - Stats: One with total

4 ER Model to Relational Model Transformation

Steps:

1. Mapping of the regular (strong) entities. Those entities are: Player, Head Coach, Team, CEO (Owner), Home Stadium, League, Awards and College. Each of these becomes its own tables with each of their attributes being columns in the table.

2. Mapping of the weak entities. Those entities are: Stats, Team Stats and Draft Records. For Stats and Team Stats, we chose to utilise the 1:1 mapping rule and merge them into one relation due to their total participation. For draft records, it became an entity, preserving its attribute, and getting a foreign key of the entity.

3. Mapping of Binary 1:1 Relations. Those relations are Player - Stats, Player - Draft Records, Team - Coach, Team - Team Stats. Since Player - Stats have total participation on each side, both entities can be merged. The same is valid for Team - Team Stats. As for the Player - Draft Records and Team - Coach relationships, the primary key of the partial participating entity is included as a foreign key for the total participating entity.

4. Mapping of Binary 1:N Relations. Those relations are Team - Player, Head Coach - Player, College - Player, CEO - Team, League - Team, Home Stadium - Team. For all of those, the primary key of the 1-side is included as a foreign key in the N-side.

5. Mapping of Binary M:N relations. There is only one M:N relation, Player - Awards. For this, a cross reference was created (named WON) and primary keys of Player and Awards were included in this 3rd relation.

6. Mapping of Multivalued attributes. There is only one multivalued attribute, Rebounds. For this, a 3rd relation (a cross reference) was created, named Play_Rebounds. The Primary Key of the owner entity (Player) was added as a foreign key.

7. There are no N-ary Relations.

8. There are no cases of Specialization or Generalization.

9. There are no Categories.

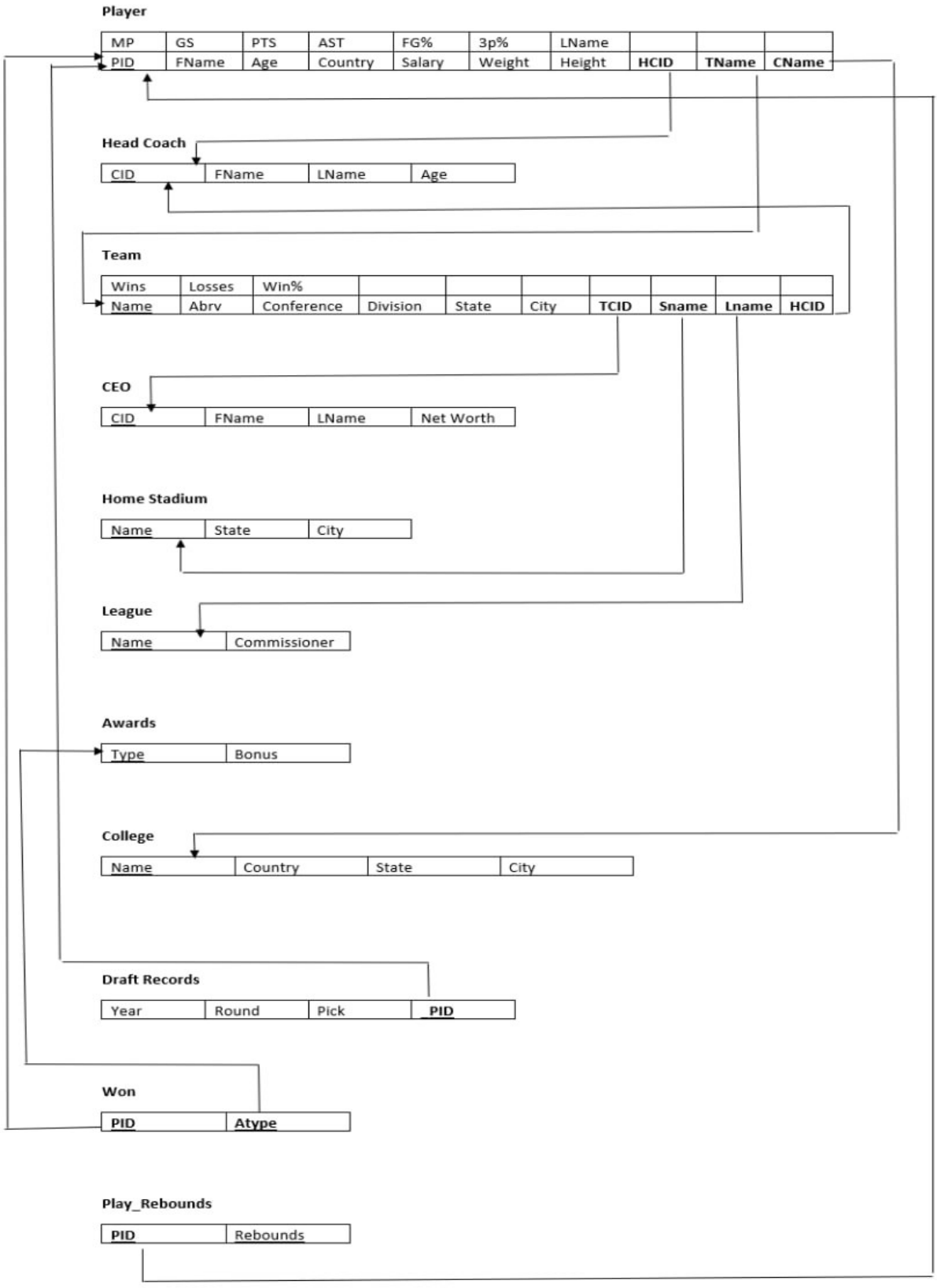

Figure 2: Relational Model Post Merge

5 Normalization

1NF:

- First normal form requires each table to have a primary key and each table to contain only atomic values. Our model satisfies these requirements.

2NF:

- Second normal form requires all non-key attributes to be functionally dependent on the Primary Key. Our model satisfies this requirement again. Every non-attribute key is functionally dependent on the primary key.

3NF:

- Third normal form is about removing transitive dependencies. It is done through decomposition. In the CEO table net_worth is dependent on name rather than CID, which is the primary key. That could be solved by decomposition. Creating a new table CEO_WEALTH that would have CEO ID as primary key and net worth, and leaving just CID and CEO name in the CEO table would solve the problem and model would satisfy the 3NF.

Boyce-Codd NF:

- The model does not violate BCNF rules of dependencies, so it satisfies BCNF.

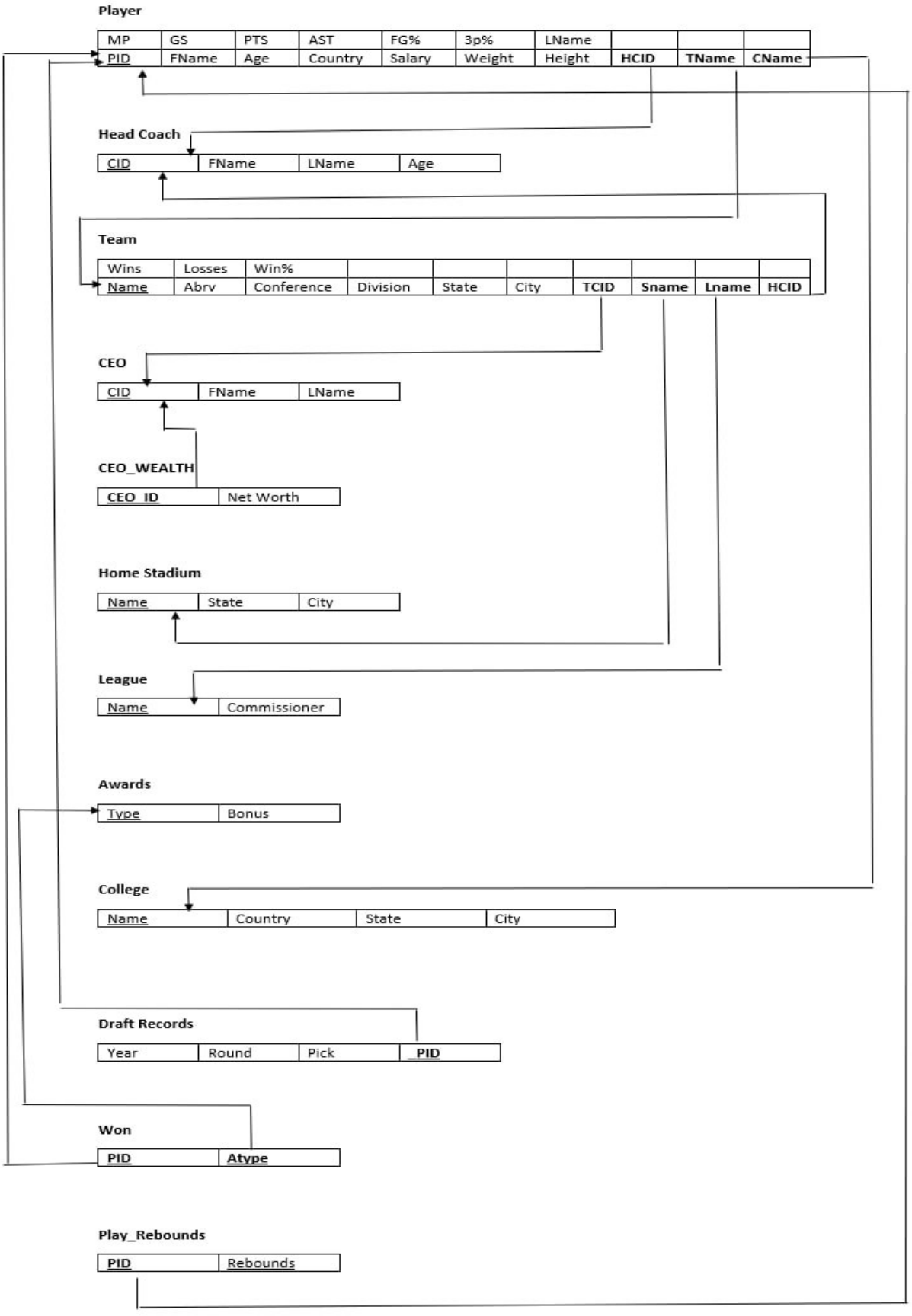

Figure 3: Normalized Relational Model

6 References

"2021–22 NBA Season." Wikipedia, Wikimedia Foundation, 3 Feb. 2023, https://en.wikipedia.org/wiki/2021%E2%80%9322_NBA_season.

Büyüknacar, Muhammet Ali. \2021–22 NBA Season Active NBA Players." Kaggle, 29 Sept. 2021, https://www.kaggle.com/datasets/buyuknacar/202122-nba-season-active-nba-players?select=active_players_2.csv.

Cirtautas, Justinas. \NBA Players." Kaggle, 6 Aug. 2022, https://www.kaggle.com/datasets/justinas/nba-players-data.

\NBA History – 2021 Awards." ESPN, ESPN Internet Ventures, http://www.espn.com/nba/history/awards/_/year/2021.

Vinco, Vivo. \2021–2022 NBA Player Stats." Kaggle, 18 June 2022, https://www.kaggle.com/datasets/vivovinco/nba-player-stats?select=2021-2022%2BNBA%2BPlayer%2BStats%2B-%2BPlayoffs.csv.