NBA Database

This is a hand-drawn Entity-Relationship (ER) diagram for an NBA Database.

Entities and attributes:

- **SeasonStats**: ppg, rpg
  - Note (blue): * points per game · rebounds per game
- **GameStats**: points, rebounds
- **Stats** — connected to SeasonStats and GameStats via a "d" (disjoint) specialization
  - Note (red): Is this the best way to handle this?
- **Players**: weight, height, name, PlayerID, bday
- **ActivePlayers**, **RetiredPlayers** — connected to Players via "d" specialization
- **Coaches**: name, CoachID
- **Team**: nickname, teamID, year founded, active Till
- **City**: cityID
- **Stadium**: capacity, Stadium ID
- **Division**: name
- **Conference**: name
- **Season**: season ID, Years covered
- **Game**: date, game ID, attendance
- **RegularSeasonGame**, **PlayOff Game** — connected to Game via "d" specialization
- **Officials**: jerseyNum, name, Official ID

Relationships:

- Players **Has** (1:N) Stats
- Stats **Has** (N:1) ...
- **ComesFrom** (N) GameStats / SeasonStats — ComesFrom
- Players **Are Teammates** (N:M)
- Players **Coached by** (N:M) Coaches
- Players **Play On** (N:M) Team — Note (green): free agency and mid-season trades
- Players **Drafted By** (N:1) Team — date
- Players **BornIn** (N:1) City
- Players **JerseyNum** (on Play On)
- Players **RetiredOn** (N:1)
- Coaches **Coach** (N:M) Team — Note (blue): e.g. coach is fired and then hired on new team
- Team **WonBy** (1:N)
- Team **PlayedIn** (N:M) Game
- Team **From** (N:1) City — Note (purple): E.g. Lakers & Clippers
- Team **Belong To** (N:1) Conference
  - Note (red): Since a team can be part of a conference should Teams be connected as well?
- Division **Part Of** (N:N) Conference
- Division **ComesFrom** (N) Season
- Season **Part Of** (1:N) Game
- City **Located In** (N:1) Stadium
- Game **Held In** (N:1) Stadium
- Stadium **Constructed** (N:1)
- Game **Officiate** (N:M) Officials
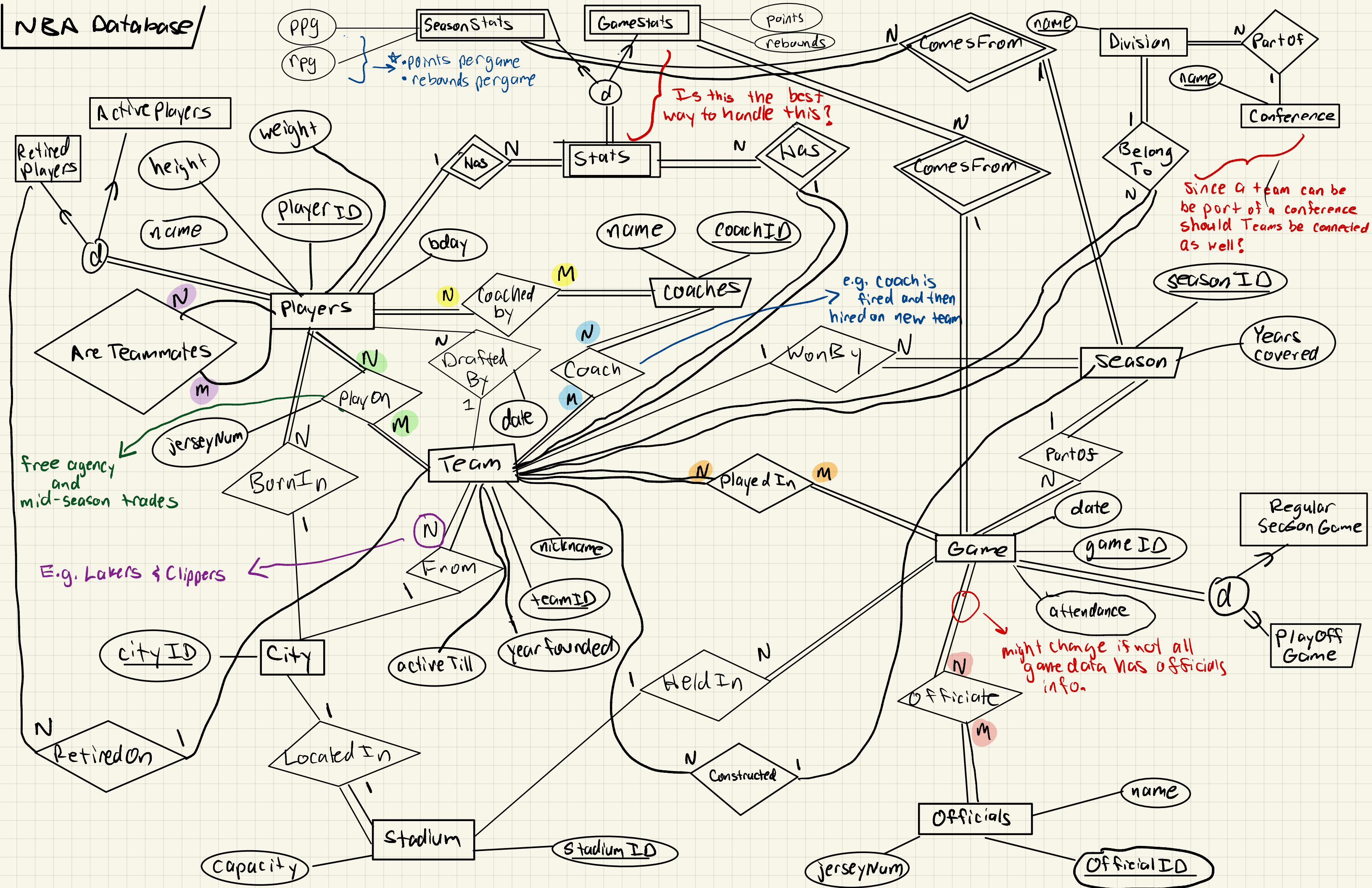 - Note (red): might change if not all game data has officials info.

- CoachedBy (Coaches/Players)
  - It is many-to-many since 1 player can get coached by multiple coaches throughout their career. One coach also coaches multiple players on a team.
  - For this particular dataset, coaches need to have at least 1 experience coaching (therefore they will need to coach at least 1 player), therefore they have total participation. Players have total participation since they must have played with at least one team, which was coached by someone
- Players subclasses
  - Players have total participation with subclasses since they are either still playing or retired. Since you can't be a retired and an active player simultaneously, active players and retired players are disjoint
- PlayOn (Players/Teams)
  - It is many-to-many since one player can play for multiple teams throughout their career and one team has multiple players on it.
  - Both have total participation because a team can't have no players on it and a player has to have played with at least one team to be in this dataset.
- BornIn (Players/City)
  - It is many-to-one from players-to-cities since one player can only be born in one city. Also, multiple players can be born in the same city.
  - Players have total participation since everyone is born in some city (unless you were born in the ocean which is a very rare case we won't even consider it!). City has partial participation since a player doesn't necessarily have to be born in it to be included in the database.
- RetiredOn (Player/Team)
  - It is many-to-one from players-to-teams since a player can only retire on one team however a team can have multiple players retire on it.
  - Players have partial participation since they may not necessarily retire from a team (e.g. they have trouble getting on a team during free agency then decide to retire there). Similarly, a team can still exist without having a single player retire on it.
- LocatedIn (City/Stadium)
  - This is 1-to-1 since a city only has one stadium in it (note: even though the Clippers are from LA, their stadium is not in LA) and a stadium can only exist in one city.
  - Cities have partial participation since they can be in the database for reasons other than a stadium being inside it (e.g. a player was born in a city where no stadium exists). Stadiums have total participation since they need to be built in some city to exist.
- From (Teams/City)
  - This is many-to-1 from teams-to-cities since one team can only be from one city, but a city can have multiple teams from it (e.g. LA Lakers and LA Clippers)
  - City has partial participation since a city can exist without being associated to a team (e.g. a player was born there). Teams have total participation since they must be associated with some city to exist.
- HeldIn (Stadium/Game)

- ○ This is 1-to-many for stadiums-to-games since one game can only be played in one stadium, but one stadium can host multiple games over time.
  - ○ a game has total participation since it must be played in some stadium to exist. Stadium has partial participation since a game might not have been played in it but can still exist in the database (e.g a brand new stadium).
- ● Officiate (Games/Officials)
  - ○ This is many-to-many since 1 game has multiple officials refereeing that game while one official can referee multiple games
  - ○ Officials have total participation since they must have officiated a game to be included in the database. Games have total participation since they need to have officials to officiate them.
- ● Game Subclasses
  - ○ Games being subclassed into regular season games or playoff games is disjoint because the regular seasons and playoffs do not occur simultaneously. Games have total participation since it MUST one of the two (i.e. there is no such things as just a game)
- ● PartOf (Game/Season)
  - ○ This is many-to-one from games-to-seasons since a game can only be included in one specific season, but one season is full of numerous games
  - ○ They both have total participation since a game MUST have occurred during some season, and a season must have had games played during it.
- ● WonBy (Team/Season)
  - ○ This is 1-to-many from teams-to-seasons since 1 team can have one multiple championships during different seasons, but one season can only have one champion from one team
  - ○ Season has total participation since it must conclude with a winner (i.e. a team must win a season for that season to end) and teams have partial participation since a team can exist in the dataset without ever winning a championship.
- ● Coach (Coach/Teams)
  - ○ This is a many-to-many relationship since one coach can coach multiple teams throughout their career and a team can hire multiple coaches (across seasons and even in the same season).
  - ○ Coaches have total participation since they must be associated with at least one time to be in the database. Also, teams have total participation since they must be coached by someone
- ● PlayedIn (Teams/Games)
  - ○ Many-to-many since one teams can play in multiple games and games must have 2 teams competing in them
  - ○ Both have total participation. Teams must be competing in some games and games wouldn't exist without teams playing in them
- ● Stats subclasses
  - ○ Stats has total participation since a stat can either be a season average or a game specific statistic
  - ○ Stats is disjoint since it can either be as season average or from a specific game
- ● ComesFrom (SeasonStats/Season)

- ○ Many-to-one since seasons stats can only come from a specific season and a season can generate multiple stats.
- ○ Season stats has total participation here since it is considered a weak entity since it can't exist without a specific season and can't be identified by itself. Also since SeasonStats is a subclass of Stats, it needs either a PlayerID or TeamID to properly identify it as well. Season has total participation since we are taking seasons that have already occurred, so it must already have stats from that season
- **ComesFrom (GameStats/Game)**
  - ○ Many-to-one since game stats can only come from a specific game and a game can generate multiple stats (e.g. for individual players and team totals).
  - ○ GameStats has total participation here since it is considered a weak entity since it can't exist without a specific game and can't be identified by itself. Also since GameStats is a subclass of Stats, it needs either a PlayerID or TeamID to properly identify it as well. Game has total participation since we are taking games that have already occurred, so it must already have stats from when it took place.
- **BelongTo (Team/Division)**
  - ○ Many-to-one since Teams can only be part of one Division and each Division contains 5 teams.
  - ○ Total participation for both since a Team by nature is associated with a division and for a division to even exist, must contain teams in it
- **PartOf (Division/Conference)**
  - ○ Many-to-one since one division is part of either Eastern or Western conference and each conference contains 15 division each
  - ○ both have total participation since a division must be associated with a conference and a conference must have teams within it to exist
- **DraftedBy (Player/Team)**
  - ○ It is many-to-one from players-to-teams since a player can only be drafted by one team however a team can draft multiple
  - ○ Players have partial participation since they may have made the NBA without being drafted (e.g. Ben Wallace, Fred Vanvleet). Similarly, a team can still exist without drafting a single player (e.g trading for picks).
- **Has (Players/Stats)**
  - ○ 1-to-many since a player can have multiple stats (e.g. season average and game stats) and one particular stat can only come from one player
  - ○ Players have total participation since everyone has stats (even players that don't play, it will just show up as 0!) and stats have total participation since it is a weak entity. That is, it can't exist without the player producing them and needs the player to properly identify who created it (as well as the game or season ID depending if it is a game stat or season stat)
- **Has (Team/Stats)**
  - ○ 1-to-many since a Team can have multiple stats (e.g. season average and game stats) and one specific stat can only come from one Team
  - ○ Teams have total participation since they will generate stats in each game and at the end of the season. Stats have total participation since it is a weak entity. That is, it can't exist without the Team producing them and needs the Team to properly

identify who created it (as well as the game or season ID depending if it is a game stat or season stat)