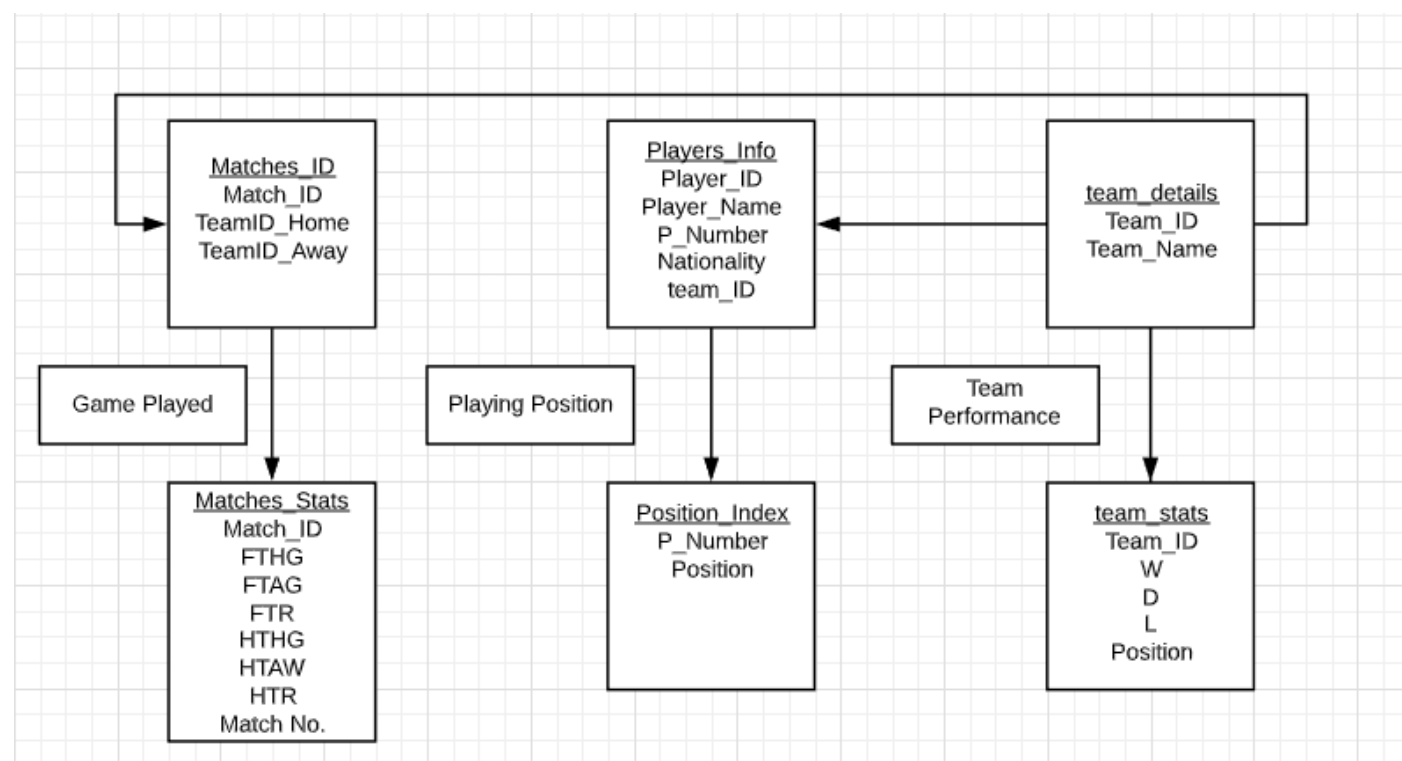# Assignment 2- Physical Data Model and Normalisation

Rohan Saji George and Yash Mahansaria (Yash/Rohan)

Domain

In continuity with the first assignment, we have decided to work on the same domain, Barclays Premier League (2018/2019) - Top 5 teams.

Conceptual Model

Attributes

1. Match_ID - Match ID for the particular match
2. TeamID_Home - Team ID of the home team for that particular match
3. TeamID_Away - Team ID of the away team for that particular match
4. FTHG- Full time Home team goals
5. FTAG- Full time Away team goals
6. FTR- Full time Result
7. HTHG- Half time Home goals
8. HTAG- Half time Away goals
9. HTR- Half time Result
10. P_Number- A designated number for the position of the player

Relationships

1. Games Played
2. Playing Position
3. Team Performance

ER Diagram
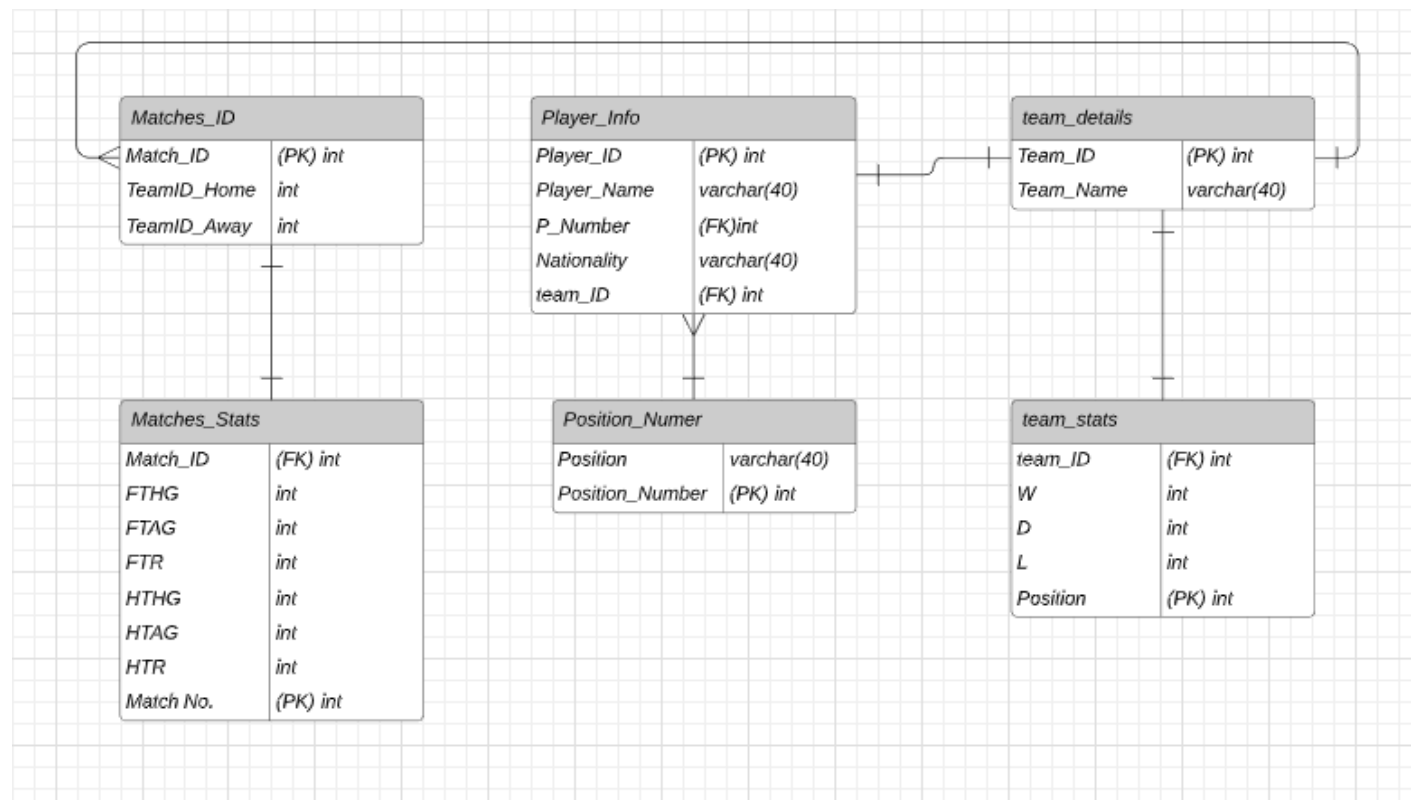
The Database is named as 2018-2019 and it has 6 tables which are mentioned in the image

Matches_ID

	Match_ID	TeamID_Home	TeamID_Away
▶	7	57	65
	11	61	57
	45	73	64
	61	61	64
	78	64	65
	99	73	65
	100	57	64
	124	73	61
	137	57	73
	154	61	65
	193	64	57
	209	65	64

Matches_Stats

	Match_ID	FTHG	FTAG	FTR	HTHG	HTAG	HTR	Match_no.
▶	11	3	2	H	2	2	D	1
	45	1	2	A	0	1	A	2
	61	1	1	D	1	0	H	3
	78	0	0	D	0	0	D	4
	99	0	1	A	0	1	A	5
	100	1	1	D	0	0	D	6
	124	3	1	H	2	0	H	7
	137	4	2	H	1	2	A	8
	154	2	0	H	1	0	H	9
	193	5	1	H	4	1	H	10
	209	2	1	H	1	0	H	11
	220	2	0	H	2	0	H	12

Player_info

	Player_ID	Player_Name	Position_Index	Nationality	Team_ID
▶	65	Mateo Kovačić	3	Croatia	61
	68	Dani Ceballos	3	Spain	57
	137	Sokratis Papastathopoulos	4	Greece	57
	145	Christian Pulisic	3	United States	61
	600	Mattéo Guendouzi	3	France	57
	1795	Alisson	5	Brazil	64
	2000	João Cancelo	4	Portugal	65
	2094	Jorginho	3	Italy	61
	2260	Lucas Torreira	3	Uruguay	57
	3141	Emiliano Martínez	5	Argentina	57
	3174	Bernd Leno	5	Germany	57
	3177	Antonio Rüdiger	4	Germany	61

Position_Number

	Position_Number	Position
▶	0	A.Manager
	1	Manager
	2	Attacker
	3	Midfielder
	4	Defender
	5	Goalkeeper

team_Details

	Team_ID	Team_Name
▶	57	Arsenal
	61	Chelsea
	64	Liverppol
	65	Manchester City
	73	Tottenham

team_stats

	Team_ID	W	D	L	Position
▶	57	21	7	10	5
	61	21	9	8	3
	64	30	7	1	2
	65	32	2	4	1
	73	23	2	13	4

General Questions

1. What are the ranges, data types and format of all of the attributes in your entities?
   - Match_ID : Numeric Data of type int
   - TeamID_Home: Numeric Data of type int
   - TeamID_Away: Numeric Data of type int
   - FTHG: Numeric Data of type int
   - FTAG: Numeric Data of type int
   - FTR: Character of type varchar(1)
   - HTHG: Numeric Data of type int
   - HTAG: Numeric Data of type int
   - HTR: Character of type varchar(1)
   - Match No.: Numeric Data of type int
   - Player_ID: Numeric Data of type int
   - Player_Name: Character of type varchar(40)
   - Nationality: Character of type varchar(40)
   - team_ID: Numeric Data of type int
   - P_Number: Numeric Data of type int
   - Position: Character of type varchar(40)
   - Team_ID: Numeric Data of type int
   - Team_Name: Character of type varchar(40)
   - W/D/L/Position: Numeric Data of type int

1. When should you use an entity versus attribute?
   - If we consider a particular player in this Database, then it is an entity but his position, nationality etc. are the attributes. Example: A player 'Hugo Lloris' is an entity in the 'player_details' table but his nationality(french) and Goalkeeper Position is the attributes to the particular entry.

1. When should you use an entity or relationship, and placement of attributes?
   - An entity depicts physical existance whereas attributes helps define different aspects(column names) of that particular entity. When it comes to relationship, it helps connect and establish a situation between two different tables in the database.

1. How did you choose your keys? Which are unique?
   - A Primary key is an unique attribute which has to be Non-Null and help in uniquely identify each row in the table.
   - A Foreign Key on the otherhand helps in refering to the primary key in another table and can contain duplicate values.
   - Primary Keys are always unique.

1. Did you model hierarchies using the "ISA" design element? Why or why not?
   - Our model doesn't have any "ISA" node because of the unique attributes of the tables which a single ISA node cannot encompass.

1. Were there design alternatives? What are their tradeoffs: entity vs. attribute, entity vs. relationship, binary vs. ternary relationships?
   - Considering all the attributes having such unique aspects, we feel this design best explains this database and no design alternatives comes in mind.

1. Where are you going find real-world data populate your model?
   - Our database has three 3 sources for real-world data to populate our model
     - An API
     - Web scraping from a website
     - a csv file which we found over the internet

Questions on Normalisation

1. Are all the tables in 1NF?
   - With a primary key in each table with atomic values and no repeating groups, yes the data is in 1NF

1. Are all the tables in 2NF?
   - There is no calculated data, all the 1NF requirements are met and all partial dependencies are omitted, making all the tables 2NF

1. Are all the tables in 3NF?
   - There are no transitive dependecies and all the 2NF requirements are met making all the tables in 3NF.