INDIAN PREMIER LEAGUE (IPL)

VAIBHAV RAJ SAHNI
2K19/CO/419

INTRODUCTION

The Indian Premier League (IPL) is a professional Twenty20 Cricket league in India usually contested between March and May of every year by eight teams representing eight different cities or states in India. The league was founded by the Board of Control for Cricket in India (BCCI) in 2007 and its first season was played in 2008. Ever since then, 13 seasons of IPL have been played and its elaborate ball-by-ball data amounts to a whole database system. In this project, I will build a MySQL database using dataset taken from Kaggle.com. The dataset consists of 6 compressed comma-separated-value (*.csv) files that are downloaded from the same.

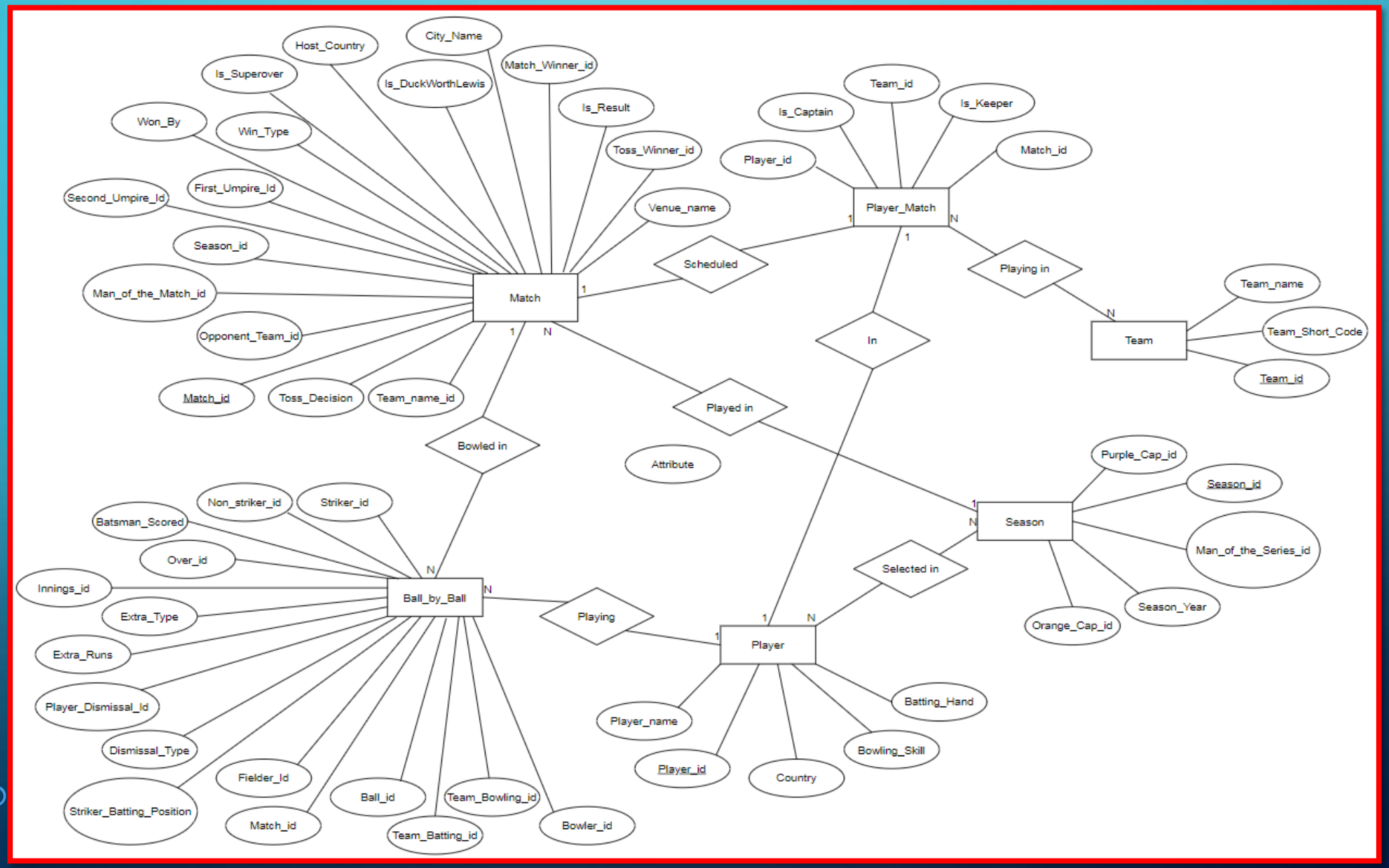
OBJECTIVES

- Understand the data in the IPL dataset
- Design a relational database and store the IPL data in it
- Model the database using an Entity-Relationship (ER) diagram
- Create its corresponding schema
- Create MySQL database
- Load data into the database
- Add primary and foreign key constraints
- Practice various SQL queries on the database

WORK DONE TILL 1ST PREVIEW

- Storing the data collected into the database
- Understanding the data
- Creating an ER diagram for the database

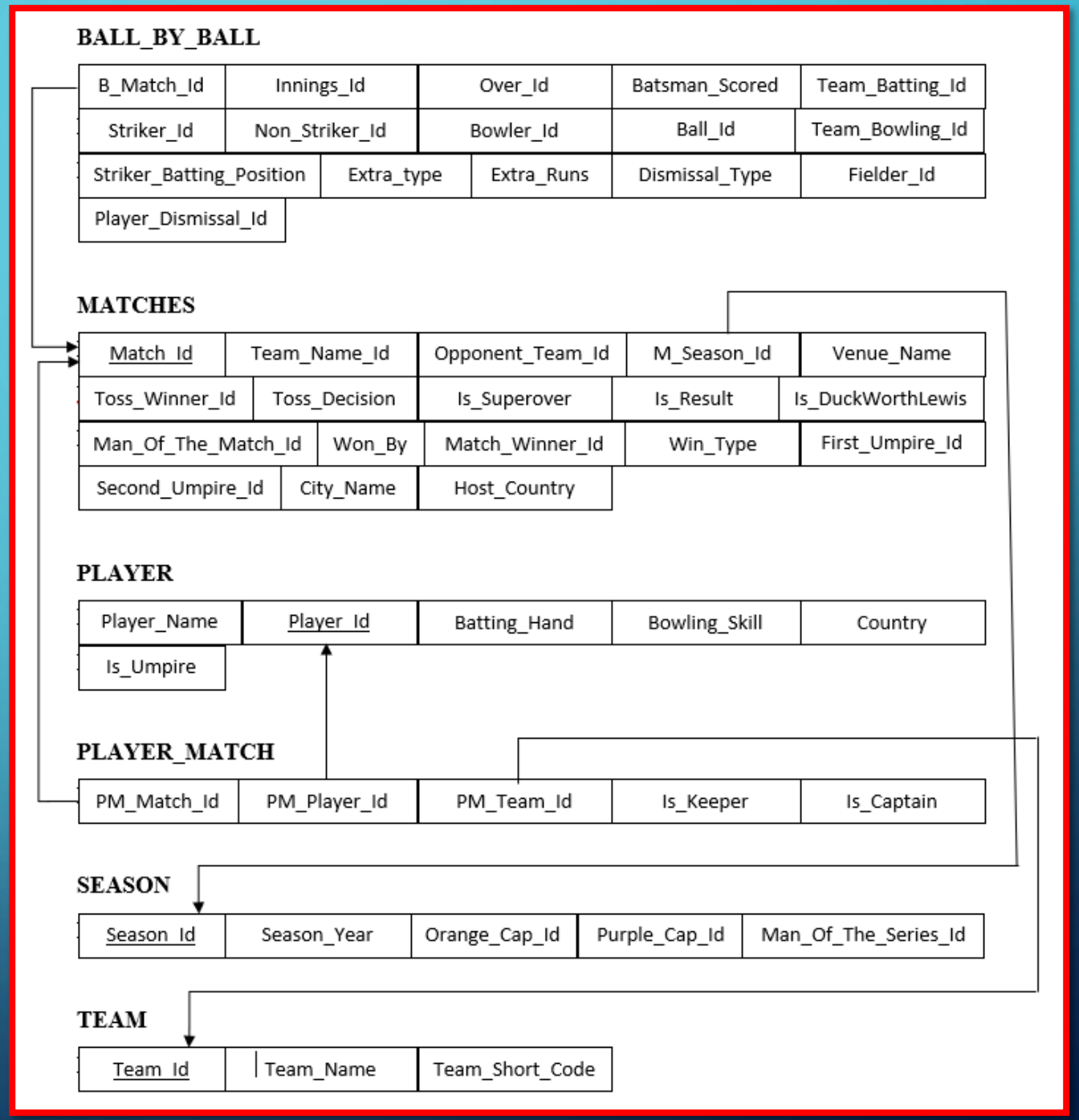
ENTITY-RELATIONSHIP DIAGRAM



WORK DONE AFTER 1ST PREVIEW

- Created a schema for the database
- Created the database in MySQL Workbench
- Imported data files using import data wizard in MySQL Workbench
- Changed all the null integer values in the data to 0 using python script
- Added Primary Key and Foreign Key constraints

SCHEMA


BALL_BY_BALL
B_Match_Id
Innings_Id
Over_Id
Batsman_Scored
Team_Batting_Id
Striker_Id
Non_Striker_Id
Bowler_Id
Ball_Id
Team_Bowling_Id
Striker_Batting_Position
Extra_type
Extra_Runs
Dismissal_Type
Fielder_Id
Player_Dismissal_Id
MATCHES
Match_Id
Team_Name_Id
Opponent_Team_Id
M_Season_Id
Venue_Name
Toss_Winner_Id
Toss_Decision
Is_Superover
Is_Result
Is_DuckWorthLewis
Man_Of_The_Match_Id
Won_By
Match_Winner_Id
Win_Type
First_Umpire_Id
Second_Umpire_Id
City_Name
Host_Country
PLAYER
Player_Name
Player_Id
Batting_Hand
Bowling_Skill
Country
Is_Umpire
PLAYER_MATCH
PM_Match_Id
PM_Player_Id
PM_Team_Id
Is_Keeper
Is_Captain
SEASON
Season_Id
Season_Year
Orange_Cap_Id
Purple_Cap_Id
Man_Of_The_Series_Id
TEAM
Team_Id
Team_Name
Team_Short_Code

PRE-PROCESSING DATA

jupyter Untitled Last Checkpoint: a day ago (autosaved) Logout

File Edit View Insert Cell Kernel Widgets Help Trusted Python 3

Run Code

In [10]:
```
import pandas as pd

# making data frame from csv file
nba = pd.read_csv("Ball_by_Ball.csv")
```

In [11]:
```
nba
```

Out[11]:

owling_Id	Striker_Id	Striker_Batting_Position	Non_Striker_Id	Bowler_Id	Batsman_Scored	Extra_Type	Extra_Runs	Player_dissimal_Id	Dissimal_Type	Fielder_Id
2	1	1	2	14	0	legbyes	1			
2	2	2	1	14	0					
2	2	2	1	14	0	wides	1			
2	2	2	1	14	0					
2	2	2	1	14	0					
...	...	...	...	...	...	...	...	...	...	...
11	369	6	434	299	2					
11	369	6	434	299	0			434	run out	183
11	140	9	369	299	0	legbyes	1			
11	369	6	140	299	1					
11	140	9	369	299	4					

Space stored in place of an integer value

BEFORE

jupyter Untitled Last Checkpoint: a day ago (autosaved)

Logout

File Edit View Insert Cell Kernel Widgets Help

Trusted Python 3

Run Code

In [16]:
```
nba["Fielder_Id"]=nba["Fielder_Id"].replace(to_replace =" ",value ="0")
```

In [17]:
```
nba
```

Out[17]:

owling_Id	Striker_Id	Striker_Batting_Position	Non_Striker_Id	Bowler_Id	Batsman_Scored	Extra_Type	Extra_Runs	Player_dissimal_Id	Dissimal_Type	Fielder_Id
2	1	1	2	14	0	legbyes	1	0		0
2	2	2	1	14	0		0	0		0
2	2	2	1	14	0	wides	1	0		0
2	2	2	1	14	0		0	0		0
2	2	2	1	14	0		0	0		0
...	...	...	...	...	...	...	...	...	...	...
11	369	6	434	299	2		0	0		0
11	369	6	434	299	0		0	434	run out	183
11	140	9	369	299	0	legbyes	1	0		0
11	369	6	140	299	1		0	0		0
11	140	9	369	299	4		0	0		0

In [18]:
```
nba.to_csv('Ball1.csv')
```

Blank space changed to value 0, which is an integer

AFTER

CREATING DATABASE

```
Create Database IPL;
Use IPL;
```

CREATING TABLES :

1. Ball_by_Ball
2. Matches
3. Player
4. Player_Match
5. Season
6. Team

```
Create Table Ball_by_Ball
(
    Match_Id integer,
    Innings_Id integer,
    Over_Id integer,
    Ball_Id integer,
    Team_Batting_Id integer,
    Team_Bowling_Id integer,
    Striker_Id integer,
    Striker_Batting_Position integer,
    Non_Striker_Id integer,
    Bowler_Id integer,
    Batsman_Scored integer,
    Extra_Type varchar(30),
    Extra_Runs integer,
    Player_Dismissal_Id integer,
    Dismissal_Type varchar(30),
    Fielder_Id integer,
    FOREIGN KEY (Match_Id) REFERENCES Matches(Match_Id)
);
```

```
Create Table Matches
(
    Match_Id integer PRIMARY KEY,
    Team_Name_Id integer,
    Opponent_Team_Id integer,
    Season_Id integer,
    Venue_Name varchar(200),
    Toss_Winner_Id integer,
    Toss_Decision varchar(15),
    Is_Superover integer,
    Is_Result integer,
    Is_DuckWorthLewis integer,
    Win_Type varchar(30),
    Won_By integer,
    Match_Winner_Id integer,
    Man_Of_The_Match_Id integer,
    First_Umpire_Id integer,
    Second_Umpire_Id integer,
    City_Name varchar(200),
    Host_Country varchar(50),
    FOREIGN KEY (Season_Id) REFERENCES Season(Season_Id)
);
```

TABLES

```
Create Table Player_Match
(
    Match_Id integer,
    Player_Id integer,
    Team_Id integer,
    Is_Keeper integer,
    Is_Captain integer,
    FOREIGN KEY (Match_Id) REFERENCES Matches(Match_Id),
    FOREIGN KEY (Player_Id) REFERENCES Player(Player_Id),
    FOREIGN KEY (Team_Id) REFERENCES Team(Team_Id)
);
```

```
Create Table Player
(
    Player_Id integer PRIMARY KEY,
    Player_Name varchar(100),
    Batting_Hand varchar(30),
    Bowling_Skill varchar(50),
    Country varchar(50),
    Is_Umpire integer
);
```

```
Create Table Season
(
    Season_Id integer PRIMARY KEY,
    Season_Year integer,
    Orange_Cap_Id integer,
    Purple_Cap_Id integer,
    Man_Of_The_Series_Id integer
);
```

```
Create Table Team
(
    Team_Id integer PRIMARY KEY,
    Team_Name varchar(60),
    Team_Short_Code varchar(4)
);
```

TABLE SHOWING SEASON DETAILS

```
Select * from Season;
```

Season_Id	Season_Year	Orange_Cap_Id	Purple_Cap_Id	Man_Of_The_Series_Id
1	2008	100	102	32
2	2009	18	61	53
3	2010	133	131	133
4	2011	162	194	162
5	2012	162	190	315
6	2013	19	71	32
7	2014	46	364	305
8	2015	187	71	334
9	2016	8	299	8

TABLE SHOWING TEAMS

```
Select * from Team;
```

Team_Id	Team_Name	Team_Short_Code
1	Kolkata Knight Riders	KKR
2	Royal Challengers Bangalore	RCB
3	Chennai Super Kings	CSK
4	Kings XI Punjab	KXIP
5	Rajasthan Royals	RR
6	Delhi Daredevils	DD
7	Mumbai Indians	MI
8	Deccan Chargers	DC
9	Kochi Tuskers Kerala	KTK
10	Pune Warriors	PW
11	Sunrisers Hyderabad	SRH
12	Rising Pune Supergiants	RPS
13	Gujarat Lions	GL

FUTURE WORK

Practice various SQL queries on the created database

REFERENCES

- https://www.kaggle.com/harsha547/indian-premier-league-csv-dataset for dataset
- App.diagrams.net for making an Entity-Relation Diagram
- https://docs.python.org/3/ for pre-processing data

THANK YOU