Description of data structure:

The Soccer_World_Cup_M data structure primarily consists of two containers: COUNTRY and STADIUM.

The COUNTRY container is a relational table that combines data from several tables within the database. It stores comprehensive details about countries, players, and their performance statistics.

Fields:

- CountryName (VARCHAR): The name of the country.
- Capital (VARCHAR): The capital city of the country.
- Population (INT): The population of the country.
- Coach (VARCHAR): The name of the country's coach.
- Player_Lname (VARCHAR): The last name of the player.
- Player_Fname (VARCHAR): The first name of the player.
- Height (DECIMAL): The height of the player.
- BirthDate (DATE): The birth date of the player.

  isCaptain (BOOLEAN): A Boolean indicating whether the player is the captain of the national team (TRUE/FALSE).
- Position (VARCHAR): The position of the player (e.g., Forward, Midfielder, etc.).
- Yellow_Cards (INT): The total number of yellow cards the player has received.
- Red_Cards (INT): The total number of red cards the player has received.
- Goals (INT): The number of goals scored by the player.
- Assists (INT): The number of assists made by the player.

Relationships:

- country: Contains information about countries.
- players: Contains details about players, including the country they are associated with.
- player_cards: Stores information on yellow and red cards received by players.
- player_assists_goals: Holds statistics for the goals and assists of each player.

The STADIUM container is a relational table that holds detailed information about match data, including the names and scores of Teams 1 and Team 2, along with the stadium name and its corresponding city.

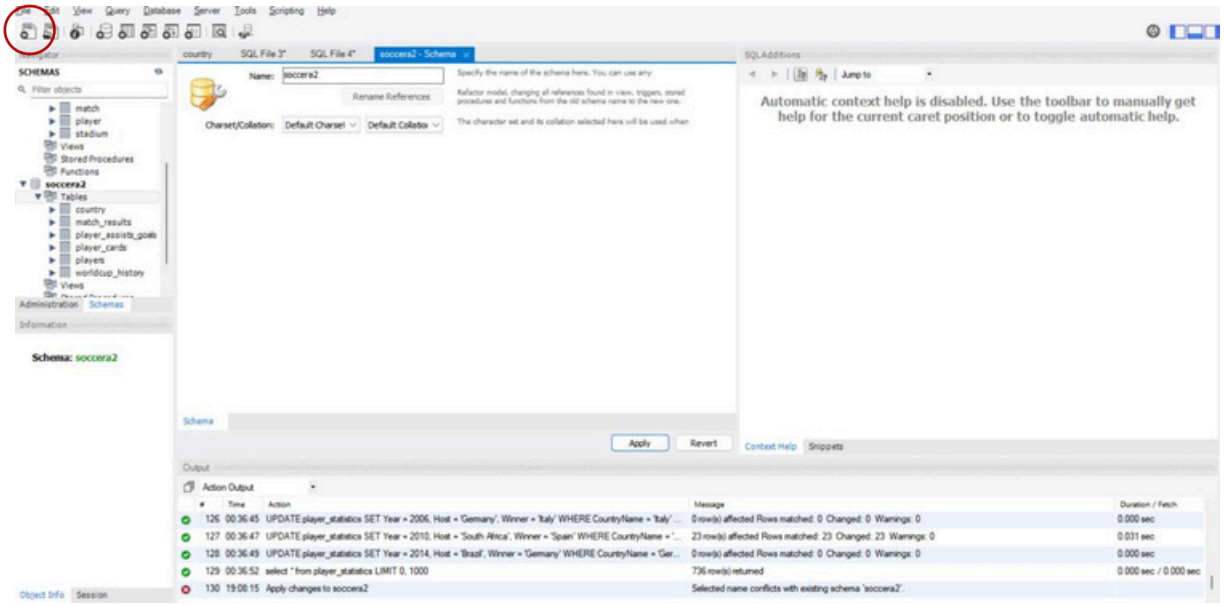
Fields:

- Match_Date (DATE): The date when the match took place.
- Team1_Name (VARCHAR): The name of the first team.
- Team2_Name (VARCHAR): The name of the second team.
- Team1_Score (INT): The score of the first team.
- Team2_Score (INT): The score of the second team.
- Stadium_Name (VARCHAR): The name of the stadium where the match was held.

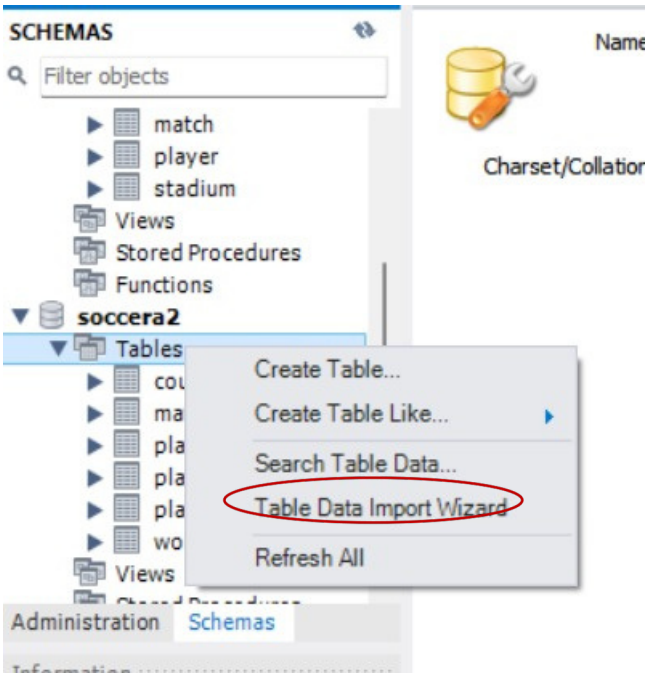
  City (VARCHAR): The city where the stadium is located.

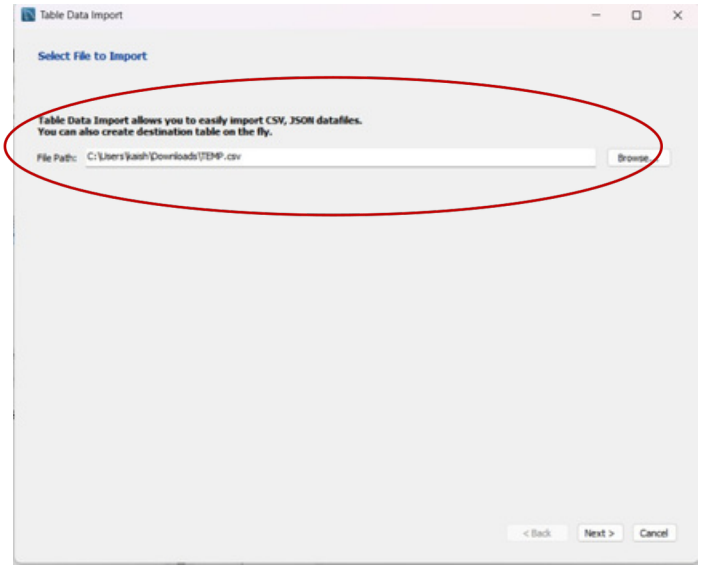
Design and implementation:

Toload datafromflatfiles, I useMySQLto accomplish this task. First, in MySQL, I create a schema named soccera2 by using the "Create Schema" button.



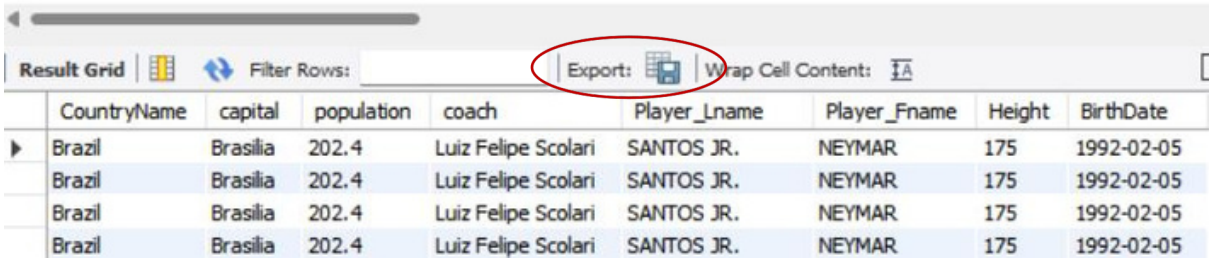
Next, I right-click on the Tables section and select the Table Data Import Wizard. Then, I load the six CSV files provided earlier—Country.csv, Players.csv, Player_Assists_Goals.csv, Player_Cards.csv, Match_Results.csv, and Worldcup_History.csv—into MySQL. These CSV files are imported as tables into the soccera2 schema from their respective file locations.





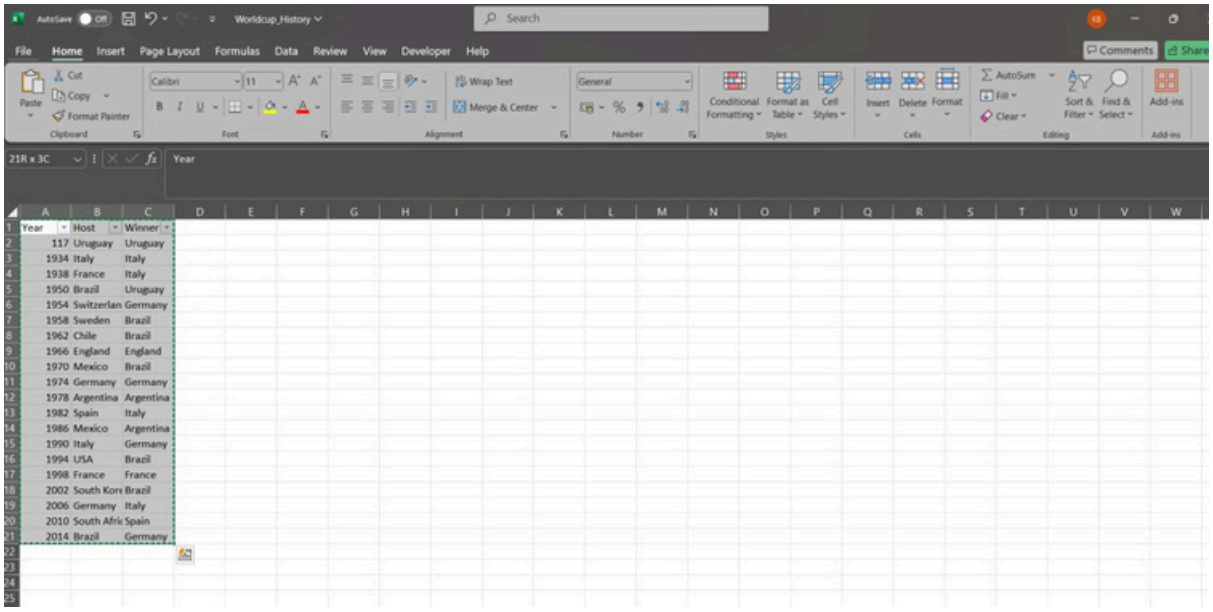
Then, we join the tables Country, Players, Player_Assists_Goals, Player_Cards, Match_Results, and Worldcup_History, and select the following columns: Cname, Capital, Population, Manager (or coach of the national team), Lname, Fname, Height, DOB, is_Captain, Position, no_Yellow_Cards, no_Red_Cards, no_Goals, and no_Assists for our COUNTRY collection.

After executing the query, we export the results to a specific location, where the data will be stored for later use when importing the files into MongoDB.

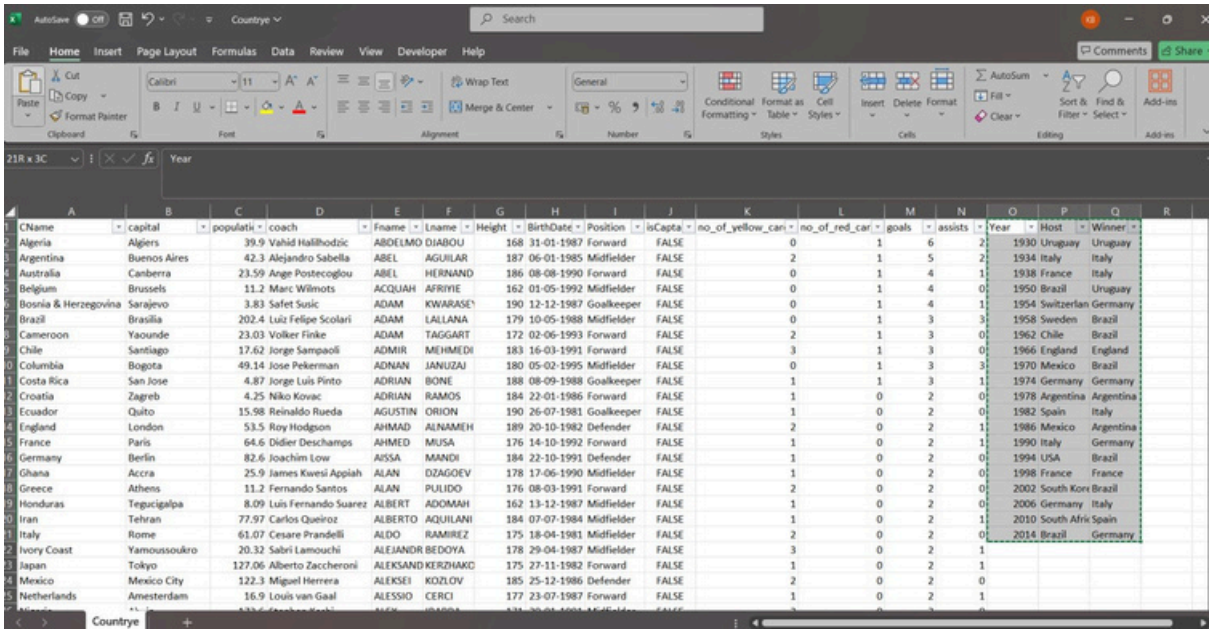


CountryName	capital	population	coach	Player_Lname	Player_Fname	Height	BirthDate
Brazil	Brasilia	202.4	Luiz Felipe Scolari	SANTOS JR.	NEYMAR	175	1992-02-05
Brazil	Brasilia	202.4	Luiz Felipe Scolari	SANTOS JR.	NEYMAR	175	1992-02-05
Brazil	Brasilia	202.4	Luiz Felipe Scolari	SANTOS JR.	NEYMAR	175	1992-02-05
Brazil	Brasilia	202.4	Luiz Felipe Scolari	SANTOS JR.	NEYMAR	175	1992-02-05

Finally, I use excel to open the exported file and add the winner, host and year by copying it from the Worldcup_History.csv and pasting it in the exported file and sav the csv file.

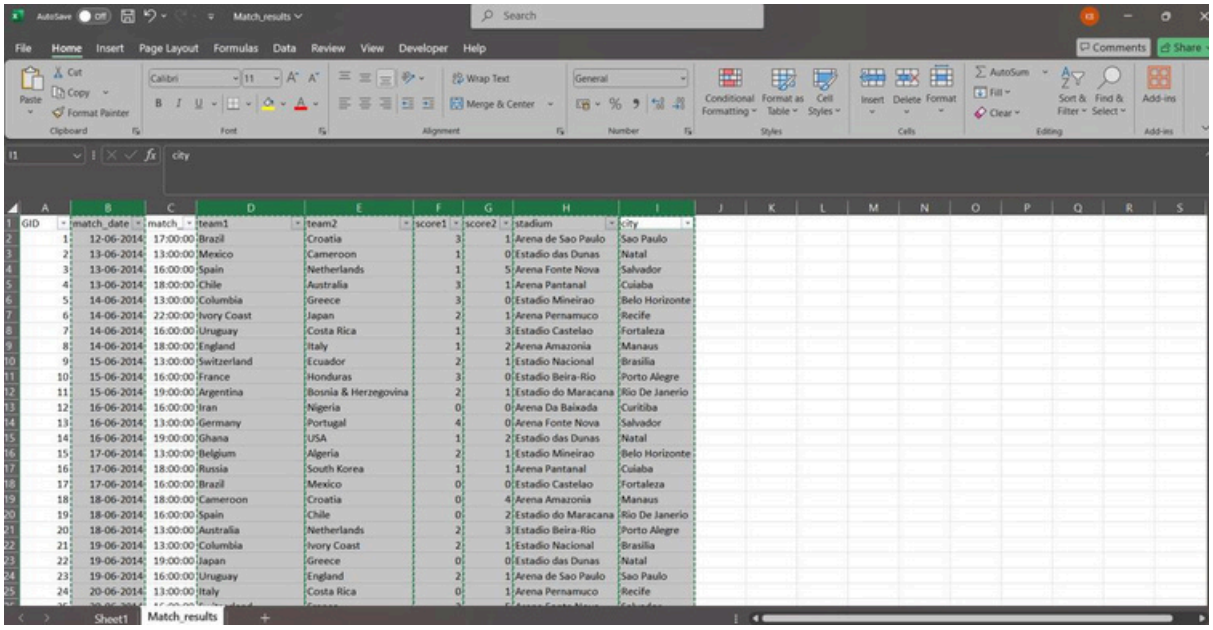


Year	Host	Winner
117	Uruguay	Uruguay
1934	Italy	Italy
1938	France	Italy
1950	Brazil	Uruguay
1954	Switzerlan	Germany
1958	Sweden	Brazil
1962	Chile	Brazil
1966	England	England
1970	Mexico	Brazil
1974	Germany	Germany
1978	Argentina	Argentina
1982	Spain	Italy
1986	Mexico	Argentina
1990	Italy	Germany
1994	USA	Brazil
1998	France	France
2002	South Kore	Brazil
2006	Germany	Italy
2010	South Afric	Spain
2014	Brazil	Germany

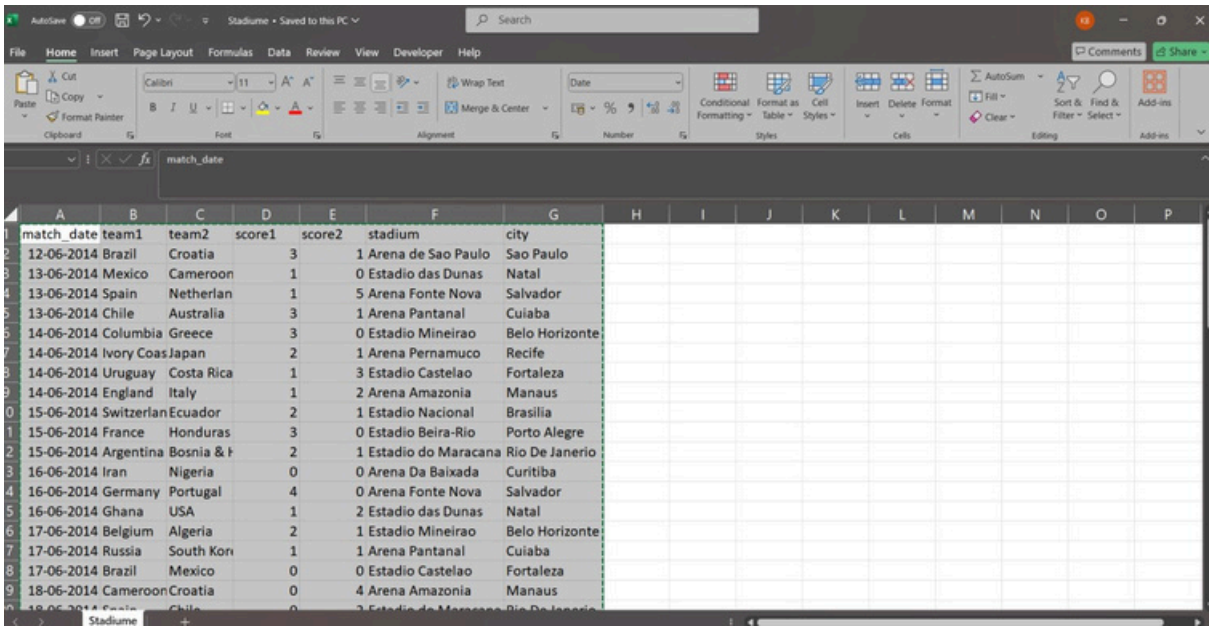


CName	capital	populati	coach	Fname	Lname	Height	BirthDate	Position	isCapta	no_of_yellow_cari	no_of_red_car	goals	assists	Year	Host	Winner
Algeria	Algiers	39.9	Vahid Halilhodzic	ABDELMO	DJABOU	168	31-01-1987	Forward	FALSE	0	1	6	2	1930	Uruguay	Uruguay
Argentina	Buenos Aires	42.3	Alejandro Sabella	ABEL	AGUILAR	187	06-01-1985	Midfielder	FALSE	2	1	5	2	1934	Italy	Italy
Australia	Canberra	23.59	Ange Postecoglou	ABEL	HERNAND	186	08-08-1990	Forward	FALSE	0	1	4	1	1938	France	Italy
Belgium	Brussels	11.2	Marc Wilmots	ACQUAH	AFRIYIE	162	01-05-1992	Midfielder	FALSE	0	1	4	0	1950	Brazil	Uruguay
Bosnia & Herzegovina	Sarajevo	3.83	Safet Susic	ADAM	KWARASE	190	12-12-1987	Goalkeeper	FALSE	0	1	4	1	1954	Switzerlan	Germany
Brazil	Brasilia	202.4	Luiz Felipe Scolari	ADAM	LALLANA	179	10-05-1988	Midfielder	FALSE	0	1	3	3	1958	Sweden	Brazil
Cameroon	Yaounde	23.03	Volker Finke	ADAM	TAGGART	172	02-06-1993	Forward	FALSE	2	1	3	0	1962	Chile	Brazil
Chile	Santiago	17.62	Jorge Sampaoli	ADMIR	MEHMEDI	183	16-03-1991	Forward	FALSE	3	1	3	0	1966	England	England
Columbia	Bogota	49.14	Jose Pekerman	ADNAN	JANUZAJ	180	05-02-1995	Midfielder	FALSE	0	1	3	3	1970	Mexico	Brazil
Costa Rica	San Jose	4.87	Jorge Luis Pinto	ADRIAN	BONE	188	08-09-1988	Goalkeeper	FALSE	1	1	3	1	1974	Germany	Germany
Croatia	Zagreb	4.25	Niko Kovac	ADRIAN	RAMOS	184	22-01-1986	Forward	FALSE	1	0	2	0	1978	Argentina	Argentina
Ecuador	Quito	15.98	Reinaldo Rueda	AGUSTIN	ORION	190	26-07-1981	Goalkeeper	FALSE	1	0	2	0	1982	Spain	Italy
England	London	53.5	Roy Hodgson	AHMAD	ALNAMEH	189	20-10-1982	Defender	FALSE	2	0	2	1	1986	Mexico	Argentina
France	Paris	64.6	Didier Deschamps	AHMED	MUSA	176	14-10-1992	Forward	FALSE	1	0	2	1	1990	Italy	Germany
Germany	Berlin	82.6	Joachim Low	AISSA	MANDI	184	22-10-1991	Defender	FALSE	1	0	2	0	1994	USA	Brazil
Ghana	Accra	25.9	James Kwesi Appiah	ALAN	DZAGOEV	178	17-06-1990	Midfielder	FALSE	1	0	2	0	1998	France	France
Greece	Athens	11.2	Fernando Santos	ALAN	PULIDO	176	08-03-1991	Forward	FALSE	2	0	2	0	2002	South Kore	Brazil
Honduras	Tegucigalpa	8.09	Luis Fernando Suarez	ALBERT	ADOMAH	162	13-12-1987	Midfielder	FALSE	1	0	2	0	2006	Germany	Italy
Iran	Tehran	77.97	Carlos Queiroz	ALBERTO	AQUILANI	184	07-07-1984	Midfielder	FALSE	1	0	2	1	2010	South Afric	Spain
Italy	Rome	61.07	Cesare Prandelli	ALDO	RAMIREZ	175	18-04-1981	Midfielder	FALSE	2	0	2	0	2014	Brazil	Germany
Ivory Coast	Yamoussoukro	20.32	Sabri Lamouchi	ALEJANDR	BEDOYA	178	29-04-1987	Midfielder	FALSE	3	0	2	1			
Japan	Tokyo	127.06	Alberto Zaccheroni	ALEKSAND	KERZHAKO	175	27-11-1982	Forward	FALSE	1	0	2	1			
Mexico	Mexico City	122.3	Miguel Herrera	ALEKSEI	KOZLOV	185	25-12-1986	Defender	FALSE	2	0	2	0			
Netherlands	Amesterdam	16.9	Louis van Gaal	ALESSIO	CERCI	177	23-07-1987	Forward	FALSE	1	0	2	1			

Next, for the STADIUM container, we use Excel to select only the following columns from the Match_Results.csv file: Stadium, City, Team1, Team2, Team1Score, Team2Score, and Date. We then paste these selected columns into a new Excel file for further processing.



GID	match_date	match_	team1	team2	score1	score2	stadium	city
1	12-06-2014	17:00:00	Brazil	Croatia	3	1	Arena de Sao Paulo	Sao Paulo
2	13-06-2014	13:00:00	Mexico	Cameroon	1	0	Estadio das Dunas	Natal
3	13-06-2014	16:00:00	Spain	Netherlands	1	5	Arena Fonte Nova	Salvador
4	13-06-2014	18:00:00	Chile	Australia	3	1	Arena Pantanal	Cuiaba
5	14-06-2014	13:00:00	Columbia	Greece	3	0	Estadio Mineirao	Belo Horizonte
6	14-06-2014	22:00:00	Ivory Coast	Japan	2	1	Arena Pernamuco	Recife
7	14-06-2014	16:00:00	Uruguay	Costa Rica	1	3	Estadio Castelao	Fortaleza
8	14-06-2014	18:00:00	England	Italy	1	2	Arena Amazonia	Manaus
9	15-06-2014	13:00:00	Switzerland	Ecuador	2	1	Estadio Nacional	Brasilia
10	15-06-2014	16:00:00	France	Honduras	3	0	Estadio Beira-Rio	Porto Alegre
11	15-06-2014	19:00:00	Argentina	Bosnia & Herzegovina	2	1	Estadio do Maracana	Rio De Janerio
12	16-06-2014	16:00:00	Iran	Nigeria	0	0	Arena Da Baixada	Curitiba
13	16-06-2014	13:00:00	Germany	Portugal	4	0	Arena Fonte Nova	Salvador
14	16-06-2014	19:00:00	Ghana	USA	1	2	Estadio das Dunas	Natal
15	17-06-2014	13:00:00	Belgium	Algeria	2	1	Estadio Mineirao	Belo Horizonte
16	17-06-2014	18:00:00	Russia	South Korea	1	1	Arena Pantanal	Cuiaba
17	17-06-2014	16:00:00	Brazil	Mexico	0	0	Estadio Castelao	Fortaleza
18	18-06-2014	18:00:00	Cameroon	Croatia	0	4	Arena Amazonia	Manaus
19	18-06-2014	16:00:00	Spain	Chile	0	2	Estadio do Maracana	Rio De Janerio
20	18-06-2014	13:00:00	Australia	Netherlands	2	3	Estadio Beira-Rio	Porto Alegre
21	19-06-2014	13:00:00	Columbia	Ivory Coast	2	1	Estadio Nacional	Brasilia
22	19-06-2014	19:00:00	Japan	Greece	0	0	Estadio das Dunas	Natal
23	19-06-2014	16:00:00	Uruguay	England	2	1	Arena de Sao Paulo	Sao Paulo
24	20-06-2014	13:00:00	Italy	Costa Rica	0	1	Arena Pernamuco	Recife

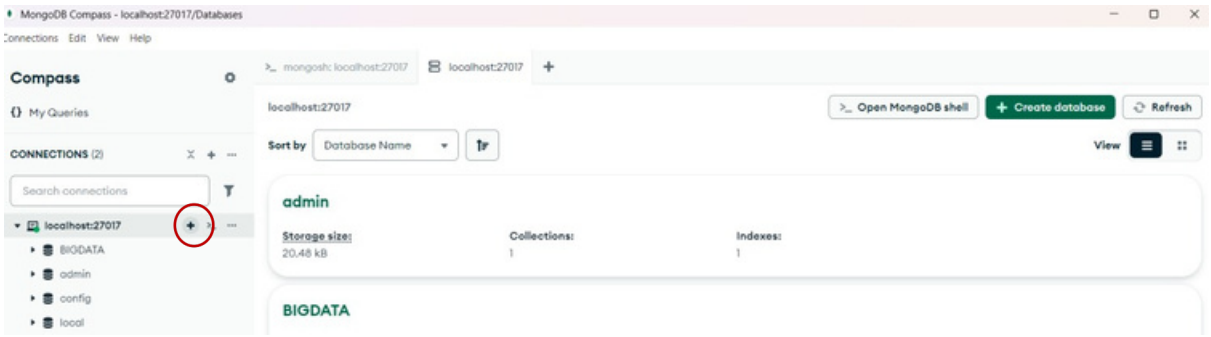


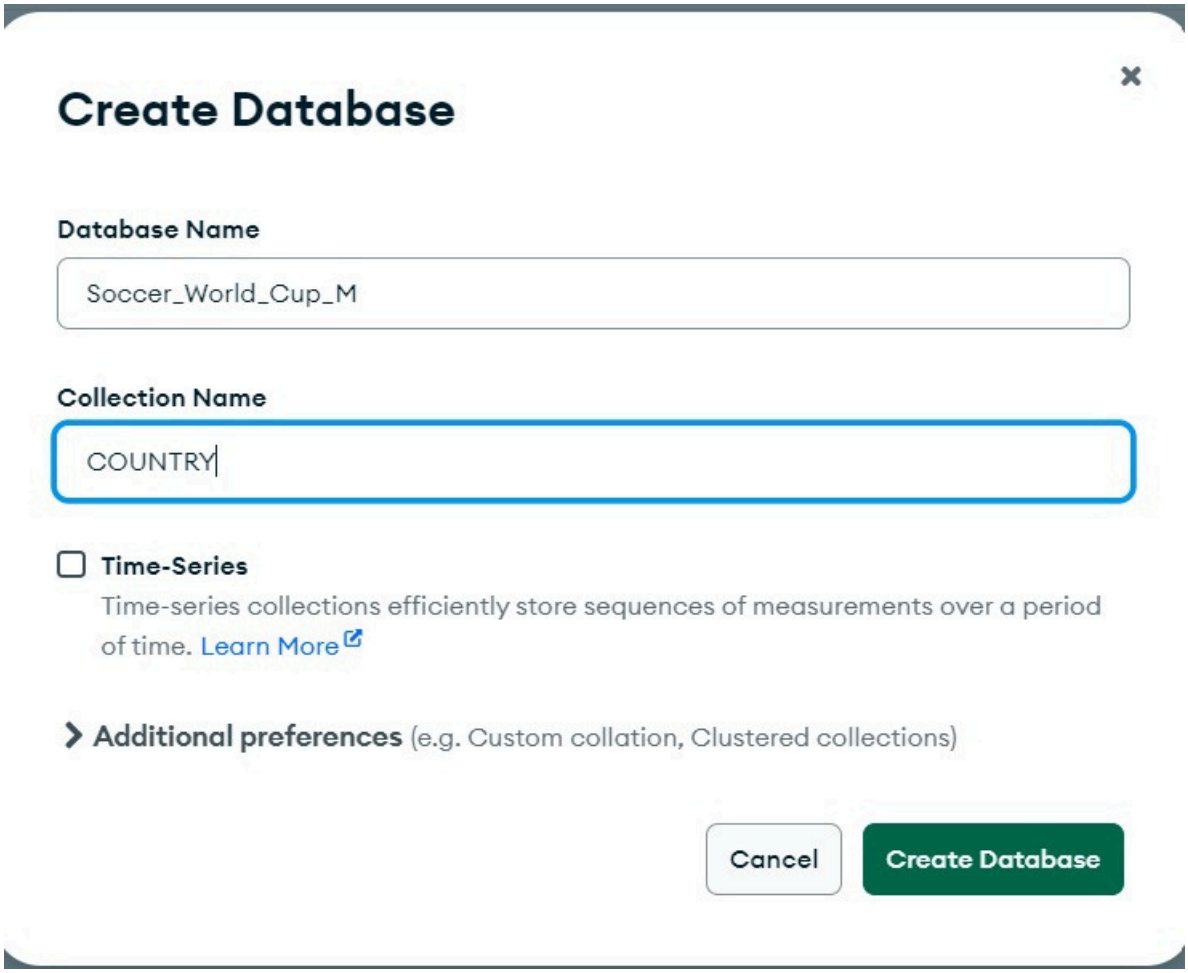
match_date	team1	team2	score1	score2	stadium	city
12-06-2014	Brazil	Croatia	3	1	Arena de Sao Paulo	Sao Paulo
13-06-2014	Mexico	Cameroon	1	0	Estadio das Dunas	Natal
13-06-2014	Spain	Netherlan	1	5	Arena Fonte Nova	Salvador
13-06-2014	Chile	Australia	3	1	Arena Pantanal	Cuiaba
14-06-2014	Columbia	Greece	3	0	Estadio Mineirao	Belo Horizonte
14-06-2014	Ivory Coas	Japan	2	1	Arena Pernamuco	Recife
14-06-2014	Uruguay	Costa Rica	1	3	Estadio Castelao	Fortaleza
14-06-2014	England	Italy	1	2	Arena Amazonia	Manaus
15-06-2014	Switzerlan	Ecuador	2	1	Estadio Nacional	Brasilia
15-06-2014	France	Honduras	3	0	Estadio Beira-Rio	Porto Alegre
15-06-2014	Argentina	Bosnia & H	2	1	Estadio do Maracana	Rio De Janerio
16-06-2014	Iran	Nigeria	0	0	Arena Da Baixada	Curitiba
16-06-2014	Germany	Portugal	4	0	Arena Fonte Nova	Salvador
16-06-2014	Ghana	USA	1	2	Estadio das Dunas	Natal
17-06-2014	Belgium	Algeria	2	1	Estadio Mineirao	Belo Horizonte
17-06-2014	Russia	South Kor	1	1	Arena Pantanal	Cuiaba
17-06-2014	Brazil	Mexico	0	0	Estadio Castelao	Fortaleza
18-06-2014	Cameroon	Croatia	0	4	Arena Amazonia	Manaus

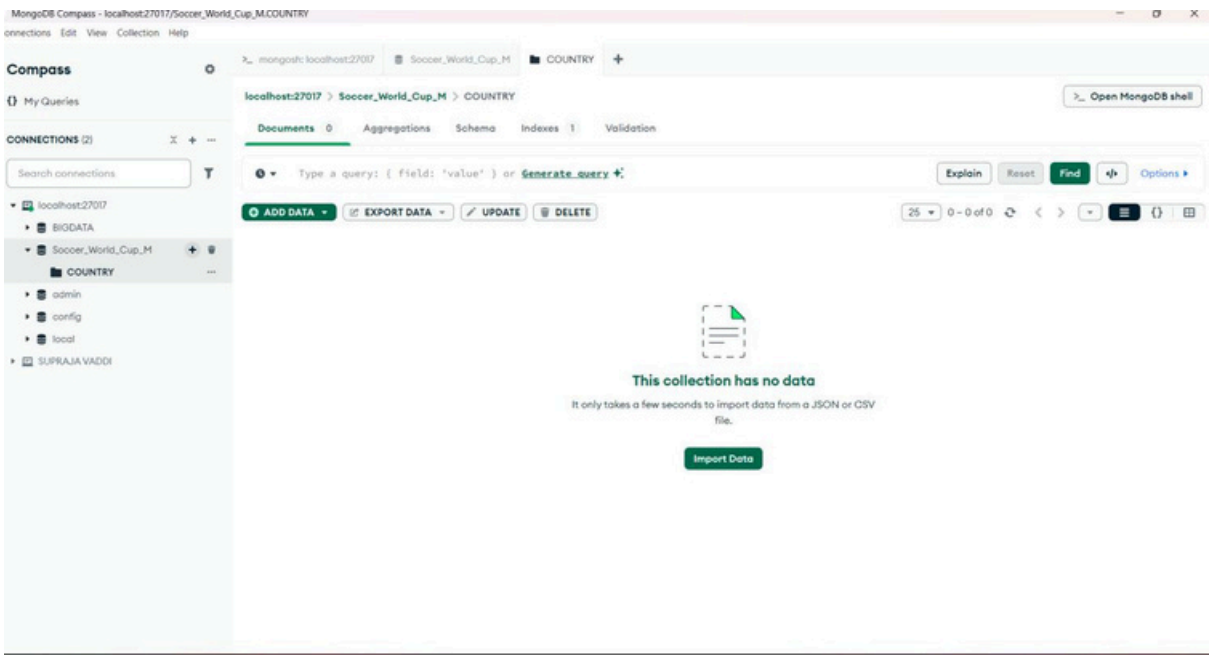
Optionally, we can also use MySQL to select only the following columns from the Match_Results table within the soccera2 schema: Stadium, City, Team1, Team2, Team1Score, Team2Score, and Date. After executing the query, we export the results to a specific location, where the data will be stored for later use when importing the files into MongoDB.

Further more we can also use python to directly get the data in Jason format. Here we first upload all the csv files and run the program. In the program we first clean the data by removing spaces and single quotes. Then we get player stats with respect to playerID which is then included in the country Json. Similarly, we only select the required column from matches_resluts.csv and save it as a Json. These JSON is then loaded into the Mongo DB System.

Next, we open MongoDB and create a new database called Soccer_World_Cup_M. Within this database, we create two collections named COUNTRY and STADIUM. After creating the collections, we import the respective CSV files into them, converting the data into JSON format to write some MongoDB queries to retrieve some of the stored documents.




Create Database
Database Name
Soccer_World_Cup_M
Collection Name
COUNTRY
Time-Series
Time-series collections efficiently store sequences of measurements over a period of time. Learn More
Additional preferences (e.g. Custom collation, Clustered collections)
Cancel
Create Database


Compass
My Queries
CONNECTIONS (2)
Search connections
localhost:27017
BIGDATA
Soccer_World_Cup_M
COUNTRY
admin
config
local
SUPRAJA VADDI
localhost:27017 > Soccer_World_Cup_M > COUNTRY
Open MongoDB shell
Documents 0
Aggregations
Schema
Indexes 1
Validation
Explain
Reset
Find
Options
ADD DATA
EXPORT DATA
UPDATE
DELETE
This collection has no data
It only takes a few seconds to import data from a JSON or CSV file.
Import Data

Source code

1. Load_country SQL file (For creation of csv of country collection)

```
-- Use the 'soccera2' database
USE soccera2;

-- Drop the 'player_statistics' table if it already exists to avoid errors
DROP TABLE IF EXISTS player_statistics;

-- Create the 'player_statistics' table by joining relevant data from multiple tables
CREATE TABLE player_statistics AS
SELECT
    -- Select country details from the 'country' table
    c.CountryName,
    c.capital,
    c.population,
    c.coach,

    -- Select player details from the 'players' table
    p.Lname AS Player_Lname,
    p.Fname AS Player_Fname,
    p.Height,
    p.BirthDate,
    p.isCaptain,
    p.Position,

    -- Use COALESCE to handle null values in the 'player_cards' table for yellow and red cards
    COALESCE(pc.no_of_yellow_cards, 0) AS no_Yellow_cards,
    COALESCE(pc.no_of_red_cards, 0) AS no_Red_cards,
```

```
   -- Use COALESCE to handle null values in the 'player_assists_goals' table for goals and assists
   COALESCE(pag.goals, 0) AS no_Goals,
   COALESCE(pag.assists, 0) AS no_Assists
FROM
   -- The 'country' table contains country details
   country c

--Join the 'players' table on the 'Country' field to match players with their respective countries
LEFT JOIN
   players p ON p.Country = c.CountryName

--Join the 'player_cards' table on the 'PID' (Player ID) to get the player's card details
LEFT JOIN
   player_cards pc ON pc.PID = p.PID

--Join the 'player_assists_goals' table on the 'PID' to get the player's goals and assists data
LEFT JOIN
   player_assists_goals pag ON pag.PID = p.PID;

--Query the newly created 'player_statistics' table to view the results
SELECT * FROM player_statistics;
```

2. Load_stadium SQL file (For creation of csv of stadium collection)

```
-- Select columns related to the match details from the match_results table
SELECT
   match_date,         -- The date when the match took place
```

```
 team1,              -- The first team playing in the match
team2,               -- The second team playing in the match
score1,              -- The score of team1 in the match
score2,              -- The score of team2 in the match
stadium,             -- The stadium where the match was held
city                 -- The city where the match was held
FROM match_results;  -- The table that contains the match results
```