Football Data Integrated System

发布 *2021-06-02*

UIC Group 16

2021 年 06 月 02 日

Contents

Project Description

1.1 Background

Football has become a hot topic with the European Championship and Champions League, and with the rise of virtual currencies, graphics cards have become a hot topic.

Due to time constraints, we chose football and graphics card these two topics as the subjects of our project at first basing on the idea of "double safety". We start these two topics togther at the same time, and if any of them encounter a problem that cannot be solved and thus cannot be carried out, we will switch to another one immediately to avoid wasting time and overdue.

In the users' feedback from JD of RTX graphics cards part, we scraped the comment time, comment star rating, comment user id and comment content of a certain RTX model by JD customers through the web crawlers. We did get these information initially, but soon Jingdong discovered that we were using crawlers and banned us. So that we finally selected the theme to be the football.

1.2 Purpose

Nowadays, more and more people are becoming fans of football, but there exists a problem: the professional data of football competitions are always charged, while the free data is lack of completeness and analysis.

Thus, we are going to provide a system collected each important football matches and provide the services of view, analysis, as well as prediction. We implement the powerful big data technology into football really will provide us some unexpected results, which may be revolutionary.

1.3 Functionalities

Our project provides platform for users to query, visualise, as well as view our rankings of most of the countries' football team, which will help them to have a clear view and intuitional understand of a country's football team. We shall make a **concept definition** here, our research object is a country's football team, thus any "team name" is equivalent to the "country name". In other words, no matter this team is a national team or a leauge team, we always use the country it belongs to as its name.

To be more specific, our application has those functionalities:

1. Points Data Visualization

   Can visually display the process of points changing for a specific team in a specific time period. An simple example is shown below.

Declaration: We only allow user to modify the database at the minimum, users are not allowed to delete any other competition event except the newest one, meanwhile, any new country user added by the adding a competition event will not be included in the calculation of points. The reasons for this kind of design is we don't want the user to spoil the database. Generally a football match's recorded is definite after the competition is over, thus the action of chaning the previous record is not need. Meanwhile, the ranking we provided is carefully selected and calculated based on big data, and any small data user input is relatively lack of reliability. If user really need to do these process, he should contact the technician to modify on the database on SQL level.

Data Description

2.1 Database Explanation

2.1.1 all_event

This table stores all the football competition events.

attribute	explanation
ID	the unique ID of competition, generated by database's trigger automatically
date	the date of the competition
home_name	the country of the home team
away_name	the country of the away team
home_score	the score of the home team
away_score	the score of the away team
tournament	the type of the competiton, e.g. FIFA World Cup, Friendly Competition, etc.
importance	the importance of the competition, will be considered in the calaculation of points as weight
result	the result of the competition, 0 - lose, 0.5 - draw, 1 - win
league_mark	the mark for identifying the affiliation of the team, 0 - national team, 1 - league team

2.1.2 world_country & league_country & all_country

These tables record the general information to evaluate a football team.

world_country table only records the national teams (whose event's league_mark=0), league_country only records the leauge teams (whose event's league_mark=1), while all_country records all the teams.

attribute	explanation
NAME	the name of the country
points	calculated points for the country
wins	the number of the winning competition
draws	the number of the competition that both teams have the same score
losses	teh number of the losing competition

2.1.3 std_uefa_ranking

This table stores the ranking of football teams by the UEFA.

attribute	explanation
Pos	the ranking of the team by this year's points
Country	the country of the football team
team_code	the short code of the country
16/17	the points of the team in 2016~2017
17/18	the points of the team in 2017~2018
18/19	the points of the team in 2018~2019
19/20	the points of the team in 2019~2020
20/21	the points of the team in 2020~2021
Pts	the total points of the team among the five years (2016~2021)

2.1.4 std_fifa_ranking

This table stores the ranking of football teams by the FIFA.

attribute	explanation
RNK	the ranking of the team by this year's points
TEAM	the country of the football team
TOTAL_POINTS	this year's points of the team
PREVIOUS_POINTS	last year's points of the team

2.2 Data Collection

We collected two dataset from Kaggle and crawled some data from the FIFA and UEFA website.

2.2.1 Dataset from Kaggle

1. Domestic Football results from 1888 to 2019: *https://www.kaggle.com/schochastics/domestic-football-results-from-1888-to-2019*

2. International football results from 1872 to 2020: *https://www.kaggle.com/martj42/international-football-results-from-1872-to-2017*

2.2.2 Crawled Data

1. FIFA World Ranking: *https://www.fifa.com/fifa-world-ranking/ranking-table/men/*

2. UEFA Country Ranking: *https://www.uefa.com/memberassociations/uefarankings/country/libraries/years/2021/*

2.3 SQL Code

For the SQL Code, please see the appendix zipped file.

Data Process

3.1 Data Preprocessing

Raw data crawled from the website will have tabs as well as tables and spaces, we use **beautifulsoap4** to ensure that we collected the data without these useless items. Meanwhile, data collected from Kaggle will have unneeded attributes and duplicated data, we use SQL statements to remove them.

3.2 Data Analysis
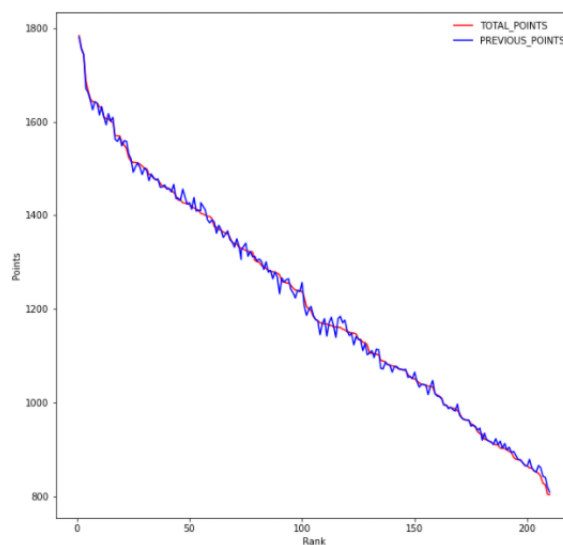
This is the plot showing the relationship between the **points** as well as the *ranking*. It's easy to discover that the changing of points and rankings are generally following a linear relationship with a slope = -1.
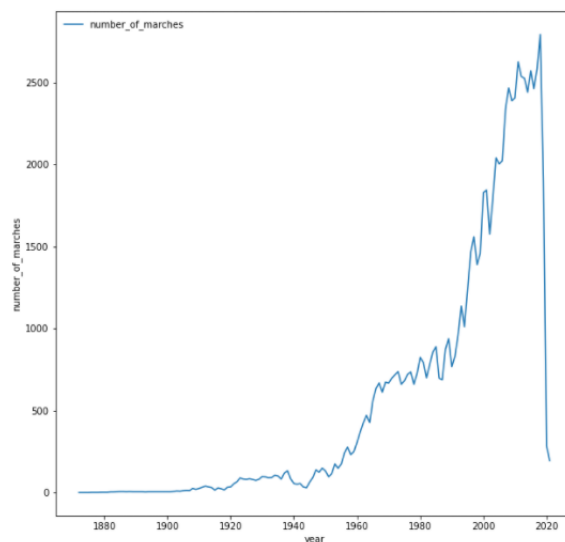
This is the plot showing the relationship between the **annual number of matches** as well as the *year*. You can discover that the annual number of matches is increasing rapidly, which shows the football area is becoming more and more prosperous.
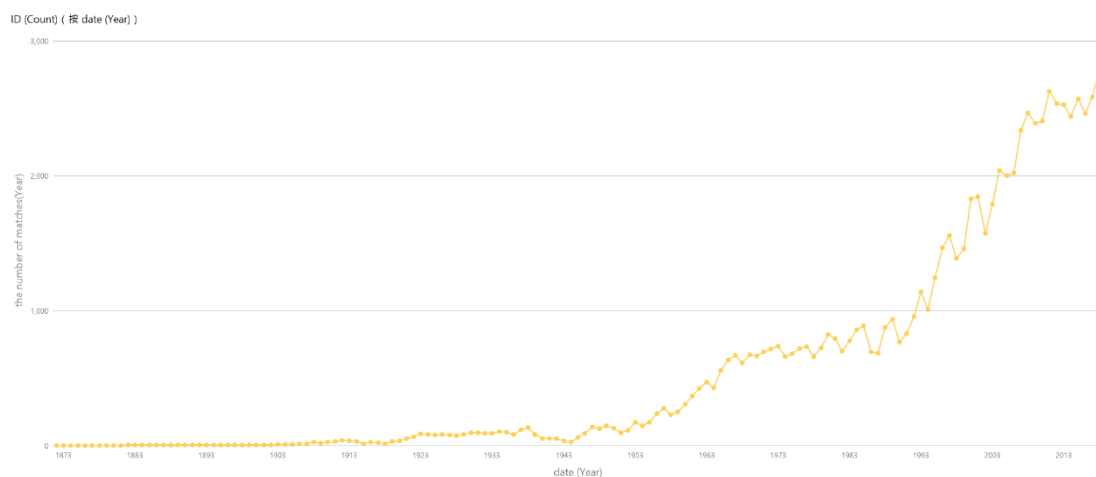
This figure reflects the **average number of football matches** from 1873 to now. It can be seen from the figure that with the increase of years, the number of football matches is gradually increasing, and after 1990, it shows a blowout growth, reflecting the continuous expansion of the scale of football.
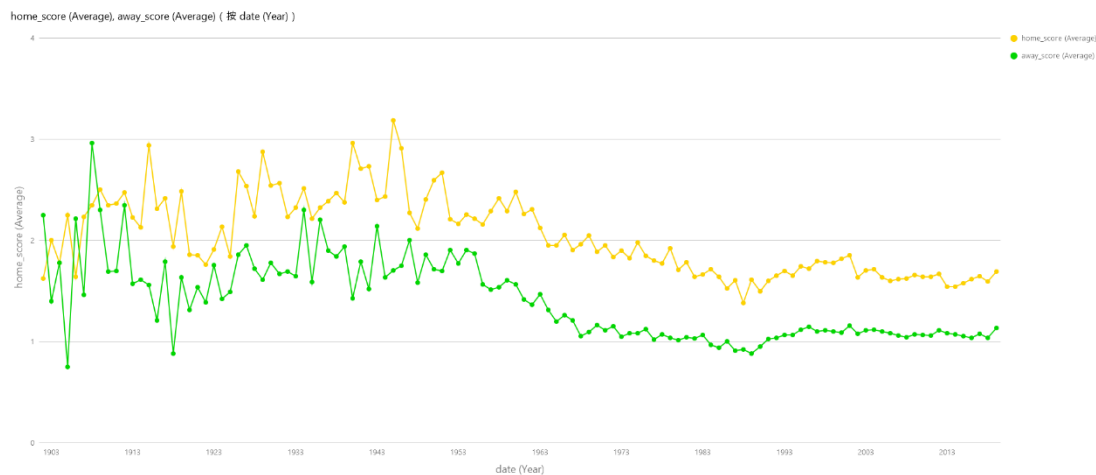
This graph shows the **number of goals scored per game**, for home team and away team, from the beginning of the 20th century to the present, sorted by year on the horizontal axis, goals scored per game on the vertical axis, goals scored at home in yellow and goals scored away in green. This graph shows two points:

1. With the development of football, home advantage gradually emerges. In each match, about 0.5 more goals are scored at home than away games;

2. With the development of football, the average goal per game shows a trend of gradual decline, which also reflects the tactical development of modern football from the side, making the defense of each team better and better.

CHAPTER 4

Database Theory

4.1 ER Diagram
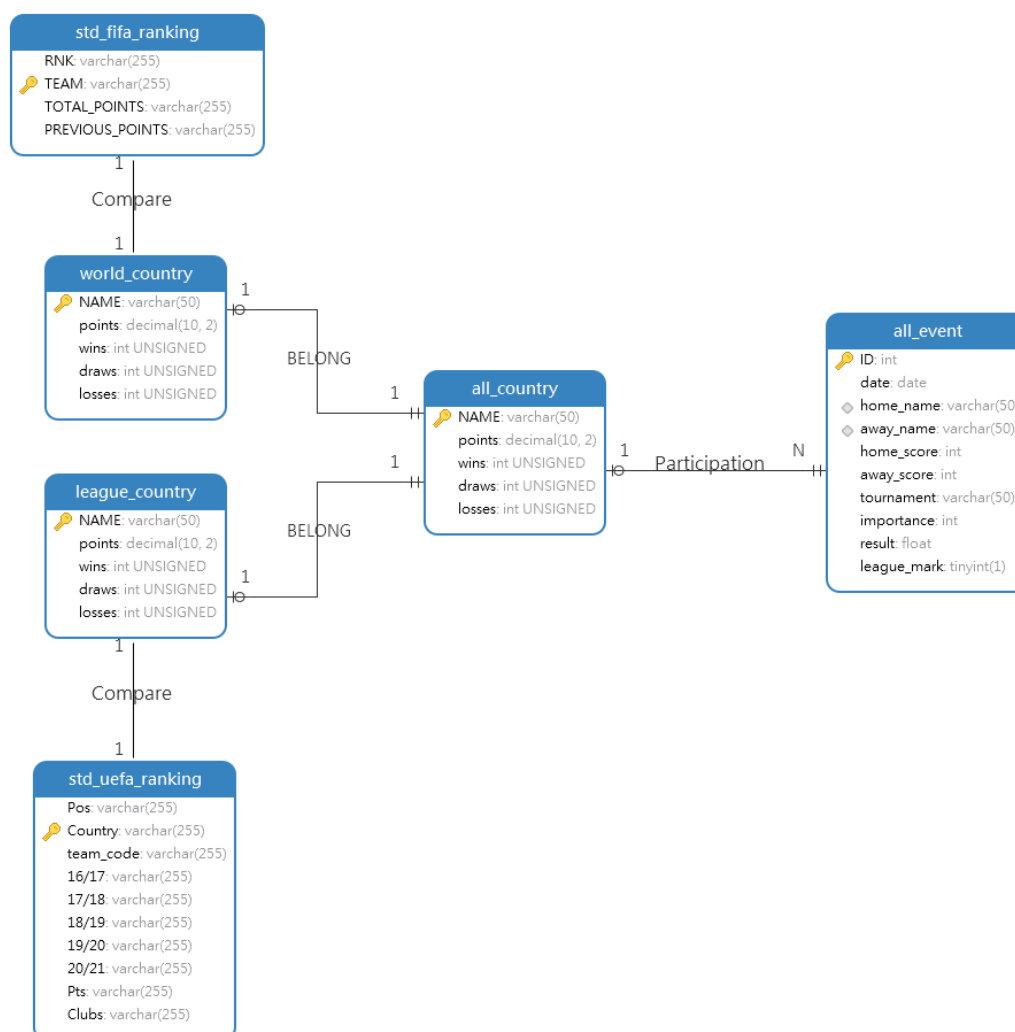
4.2 Functional Dependencies

Table: all_country
NAME→points, wins, draws, losses

Table: all_event
ID→date, home_name, away_name, home_score, away_score, tournament, importance, result, league_mark
tournament←→importance
home_score, away_score→result

Table：world_country
NAME→points, wins, draws, losses

Table：league_country
NAME→points, wins, draws, losses

Table：std_fifa_ranking
TEAM→TOTAL_POINTS, PREVIOUS_POINTS
TOTAL_POINTS→RNK

Table: std_uefa_ranking
Country←→Pos
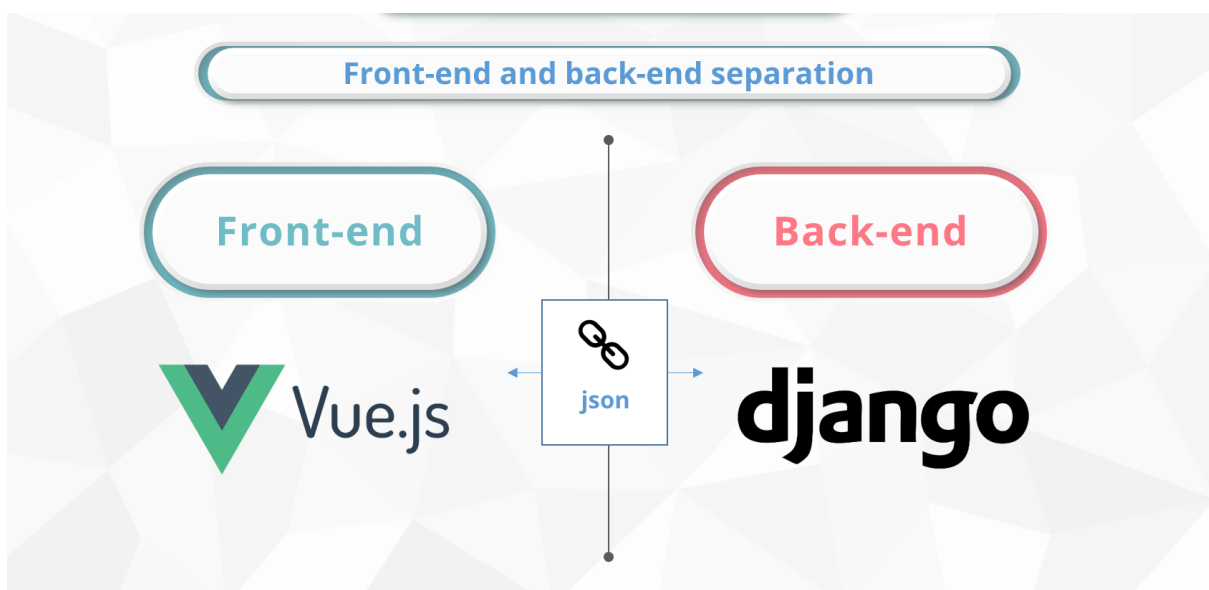Country←→team_code
Country →16/17,17/18,18/19,19/20,20/21,Pts,Clubs
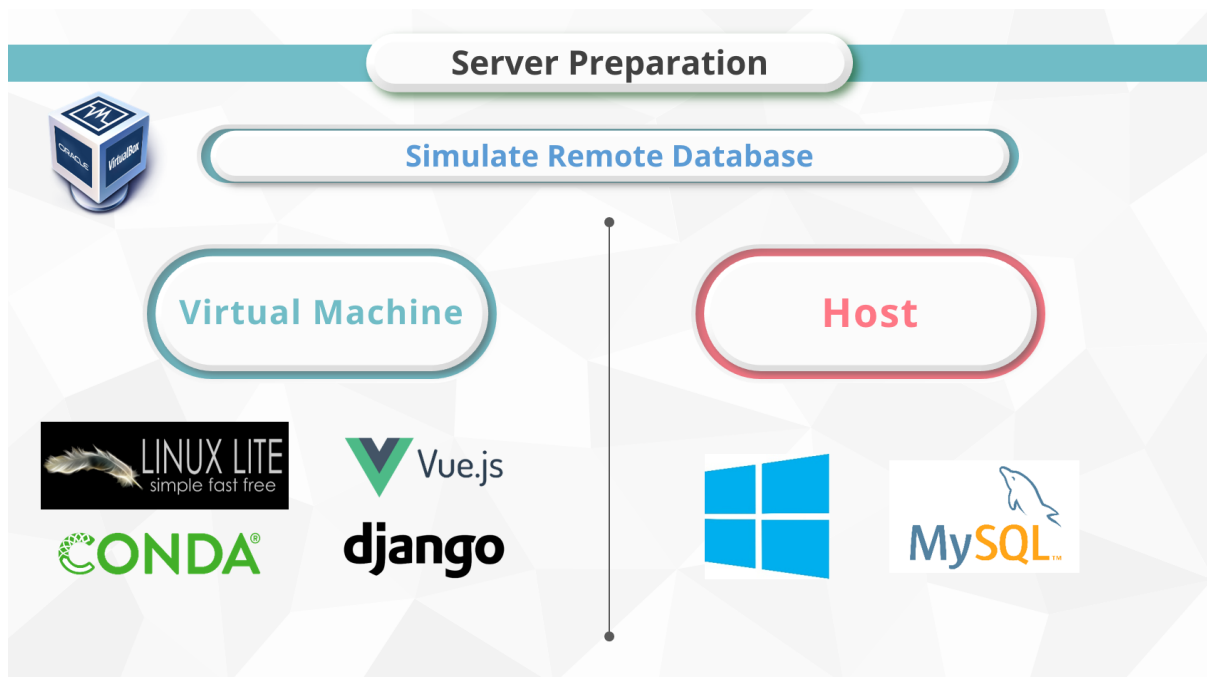
Website Application

5.1 Development Architecture

We use the "Front-end and back-end separation" development mode, for the front-end we use **vue.js**, and for the back-end we use **django**.

5.2 Server Configuration

We use virtual machine to simulate the case that the database server and the application server is not on the same host machine.

5.3 feature implementation

Feature implementation is introduced at the first chapter "Project Description".

CHAPTER 6

Difficulty

Time, is the only difficulty. If we got more time, we will absolutely make it better!