

Fall 2022 - CS 673: Scalable Databases

End Semester Project

by

Pushpak Atey, Karan Ksheersagar, Darshana Choudhary

Project Title: FIFA World cup stats using Neo4j

PROBLEM STATEMENT:

- The Data is related to FIFA World cup and its Statistics of top 15 players.
- The Problem statement is based on performing Graphical Representation and its co-relation with various entities.
- Created database of Players, Coaches, Country and defined unique relationships between them.

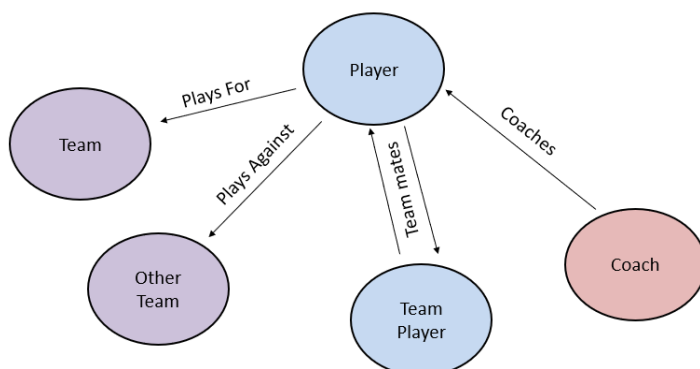
INTRODUCTION:

- We have created a graphical data of FIFA World Cup and its relationship of a player with other teammates, relationship of coaches with players and players with their respective country.
- We have taken a data of 6 countries which have been qualified through playoffs and their top 15 players with 6 coaches and created individual nodes.

WHY GRAPH DATABASE?

- The ability to add new nodes and relationships between nodes to a graph database makes it reliable for our dataset.
- Graph databases respond to complex queries more quickly than relational databases do.
- We have multiple relationship in our database so in order to display it without any constraint we have used neo4j.
- As per future aspects in our database we can create multiple relationship with ease and visualization is better.

DATABASE MODEL:



CREATE DATABASE:


- Created dataset of 15 players of FIFA World Cup and added their properties as age, number, height, weight to its nodes.
- Created 6 nodes of coaches with respect to their players.

```
1 CREATE
2 (messi:PLAYER{name:"Lionel Messi", age: 35, number: 10, height: 1.7, weight: 68}),
3 (maria:PLAYER{name:"Angel Di Maria", age: 34, number: 11, height: 1.8, weight: 70}),
4 (alvarez:PLAYER{name:"Julian Alvarez", age: 22, number: 9, height: 1.7, weight: 71}),
5 (ronaldo:PLAYER{name:"Cristiano Ronaldo", age: 37, number: 7, height: 1.87, weight: 83}),
6 (bruno:PLAYER{name:"Bruno Fernandes", age: 28, number:8, height: 1.79, weight: 69}),
7 (neymar:PLAYER{name:"Neymar Jr", age: 30, number: 10, height: 1.75, weight: 68}),
8 (vini:PLAYER{name:"Vinicius Jr", age: 22, number:20, height: 1.76, weight: 73}),
9 (silva:PLAYER{name:"thiago Silva", age: 38, number: 3, height: 1.83, weight: 76}),
10 (mbappe:PLAYER{name:"Kylia n Mbappe", age: 23, number: 10, height: 1.78, weight: 75}),
11 (griezmann:PLAYER{name:"Antoine Griezmann", age: 31, number: 7, height: 1.76, weight: 73}),
12 (giroud:PLAYER{name:"Olivier Giroud", age: 36, number: 9, height: 1.93, weight: 91}),
```

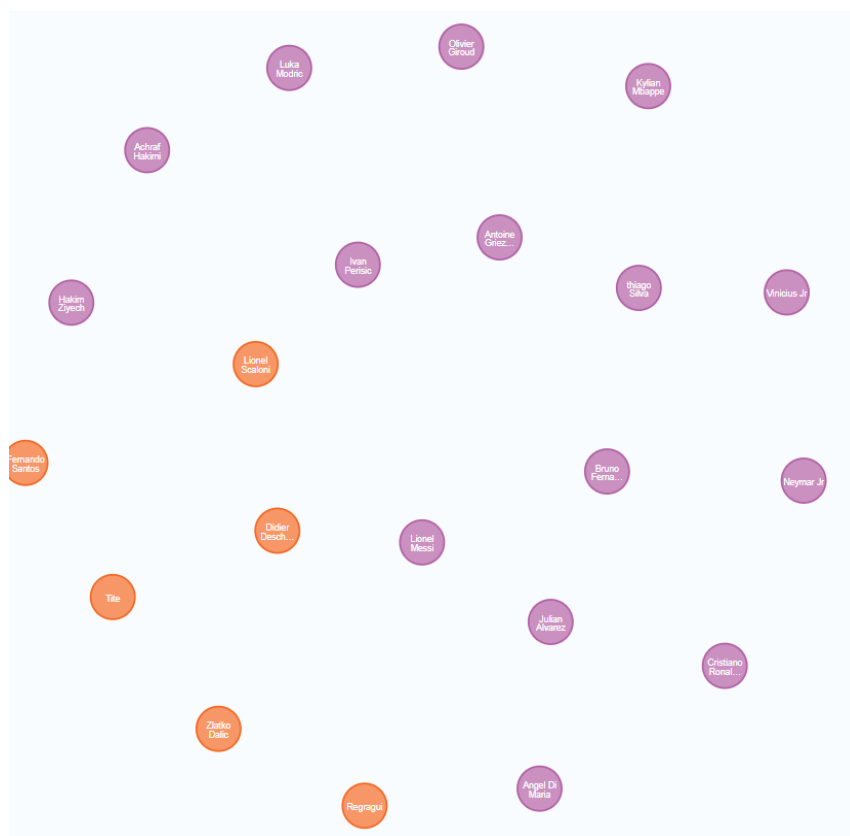
 Added 15 labels, created 15 nodes, set 75 properties, completed after 52 ms.

- Coach's nodes have been shown below:

```
1 CREATE
2 (lio:COACH{name: "Lionel Scaloni"}),
3 (santos:COACH{name: "Fernando Santos"}),
4 (tite:COACH{name: "Tite"}),
5 (didier:COACH{name: "Didier Deschamps"}),
6 (dalic:COACH{name: "Zlatko Dalic"}),
7 (regra:COACH{name: "Regragui"})
8
```

 Added 6 labels, created 6 nodes, set 6 properties, completed after 11 ms.

- Following is the graphical node representation of players and coaches.
- Node Properties of an individual player node is shown below.



Node properties ⓘ

PLAYER

<id>	7	ⓘ
age	38	ⓘ
height	1.83	ⓘ
name	thiago Silva	ⓘ
number	3	ⓘ
weight	76	ⓘ

- Created 6 nodes of different teams participated in FIFA World Cup.
- Created dual relationship of team mates between players node.

```

1 CREATE
2 (arg:TEAM{name:"Argentina"}),
3 (por:TEAM{name:"Protugal"}),
4 (fra:TEAM{name:"France"}),
5 (bra:TEAM{name:"Brazil"}),
6 (cro:TEAM{name:"Croatia"}),
7 (mor:TEAM{name:"Morocco"})

```



Added 6 labels, created 6 nodes, set 6 properties, completed after 10 ms.

```

1 CREATE
2 (messi)-[:TEAMMATES]→ (maria),
3 (messi)←[:TEAMMATES]- (maria),
4 (maria)-[:TEAMMATES]→ (alvarez),
5 (maria)←[:TEAMMATES]- (alvarez),
6 (alvarez)-[:TEAMMATES]→ (messi),
7 (alvarez)←[:TEAMMATES]- (messi),
8 (ronaldo)-[:TEAMMATES]→ (bruno),
9 (ronaldo)←[:TEAMMATES]- (bruno),
10 (neymar)-[:TEAMMATES]→ (vini),
11 (neymar)←[:TEAMMATES]- (vini),
12 (vini)-[:TEAMMATES]→ (silva),

```



Created 15 nodes, created 24 relationships, completed after 34 ms.

- Created Relationship between players, teams (node) and gave properties.
- We have created property named 'appearance'; it shows players appearance for the respected team.

```

1 CREATE
2 (messi)-[:PLAYS {appearances: 172}]]→ (arg),
3 (maria)-[:PLAYS {appearances: 120}]]→ (arg),
4 (alvarez)-[:PLAYS {appearances: 19}]]→ (arg),
5 (ronaldo)-[:PLAYS {appearances: 196}]]→ (por),
6 (bruno)-[:PLAYS {appearances: 53}]]→ (por),
7 (neymar)-[:PLAYS {appearances: 124}]]→ (bra),
8 (vini)-[:PLAYS {appearances: 20}]]→ (bra),
9 (silva)-[:PLAYS {appearances: 113}]]> (bra),
10 (mbappe)-[:PLAYS {appearances: 66}]]→ (fra),
11 (griezmann)-[:PLAYS {appearances: 117}]]→ (fra),
12 (giroud)-[:PLAYS {appearances: 120}]]→ (fra),
13 (modric)-[:PLAYS {appearances: 162}]]→ (cro),
14 (ivan)-[:PLAYS {appearances: 123}]]→ (cro),
15 (hakimi)-[:PLAYS {appearances: 61}]]→ (mor),
16 (ziyech)-[:PLAYS {appearances: 50}]]→ (mor)

```

- Created Relationship between players, teams (node) and gave properties of goals scored, total assists, attempts and tackle during their game.

```

1 CREATE
2 messi)-[:PLAYED_AGAINST {goal: 3, assists: 2, attempt: 6, tackle: 8}]]→ (fra),
3 (messi)-[:PLAYED_AGAINST {goal: 1, assists: 1, attempt: 3, tackle: 4}]]→ (cro),
4 (alvarez)-[:PLAYED_AGAINST {goal: 2, assists: 0, attempt: 2, tackle: 3}]]→ (cro),
5 (maria)-[:PLAYED_AGAINST {goal: 1, assists: 2, attempt: 1, tackle: 5}]]→ (fra),
6 (hakimi)-[:PLAYED_AGAINST {goal: 0, assists: 1, attempt: 2, tackle: 7}]]→ (cro),
7 (ivan)-[:PLAYED_AGAINST {goal: 0, assists: 2, attempt: 3, tackle: 4}]]→ (bra),
8 (mbappe)-[:PLAYED_AGAINST {goal: 1, assists: 2, attempt: 3, tackle: 4}]]→ (mor),
9 (mbappe)-[:PLAYED_AGAINST {goal: 3, assists: 0, attempt: 4, tackle: 3}]]→ (arg),
10 (griezmann)-[:PLAYED_AGAINST {goal: 0, assists: 0, attempt: 2, tackle: 2}]]→ (arg),
11 (maria)-[:PLAYED_AGAINST {goal: 1, assists: 0, attempt: 2, tackle: 6}]]→ (cro),
12 (ronaldo)-[:PLAYED_AGAINST {goal: 1, assists: 0, attempt: 2, tackle: 3}]]→ (mor),
13 (bruno)-[:PLAYED_AGAINST {goal: 1, assists: 1, attempt: 0, tackle: 1}]]→ (mor),
14 (neymar)-[:PLAYED_AGAINST {goal: 2, assists: 0, attempt: 4, tackle: 3}]]→ (cro),
15 (vini)-[:PLAYED_AGAINST {goal: 0, assists: 1, attempt: 1, tackle: 1}]]→ (cro),
16 (silva)-[:PLAYED_AGAINST {goal: 1, assists: 0, attempt: 4, tackle: 3}]]→ (arg),
17 (giroud)-[:PLAYED_AGAINST {goal: 2, assists: 1, attempt: 3, tackle: 2}]]→ (por),
18 (ziyech)-[:PLAYED_AGAINST {goal: 0, assists: 0, attempt: 1, tackle: 2}]]→ (por),
19 (modric)-[:PLAYED_AGAINST {goal: 3, assists: 0, attempt: 5, tackle: 7}]]→ (bra),
20 (ivan)-[:PLAYED_AGAINST {goal: 1, assists: 2, attempt: 2, tackle: 2}]]→ (bra),
21 (hakimi)-[:PLAYED_AGAINST {goal: 1, assists: 0, attempt: 2, tackle: 3}]]→ (fra),
22 (neymar)-[:PLAYED_AGAINST {goal: 2, assists: 3, attempt: 4, tackle: 5}]]→ (por),
23 (vini)-[:PLAYED_AGAINST {goal: 1, assists: 0, attempt: 2, tackle: 1}]]→ (mor),
24 (ziyech)-[:PLAYED_AGAINST {goal: 0, assists: 1, attempt: 1, tackle: 3}]]→ (fra);

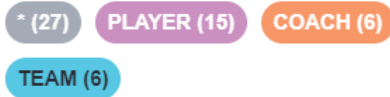
```

TRANSFORM DATABASE

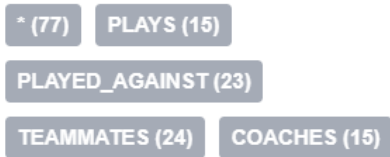
- Final Graphical Representation of created data and its relationships.
- There are 15 nodes of Players, 6 nodes and Team and Coach respectively, total relationships created are 77.

Overview

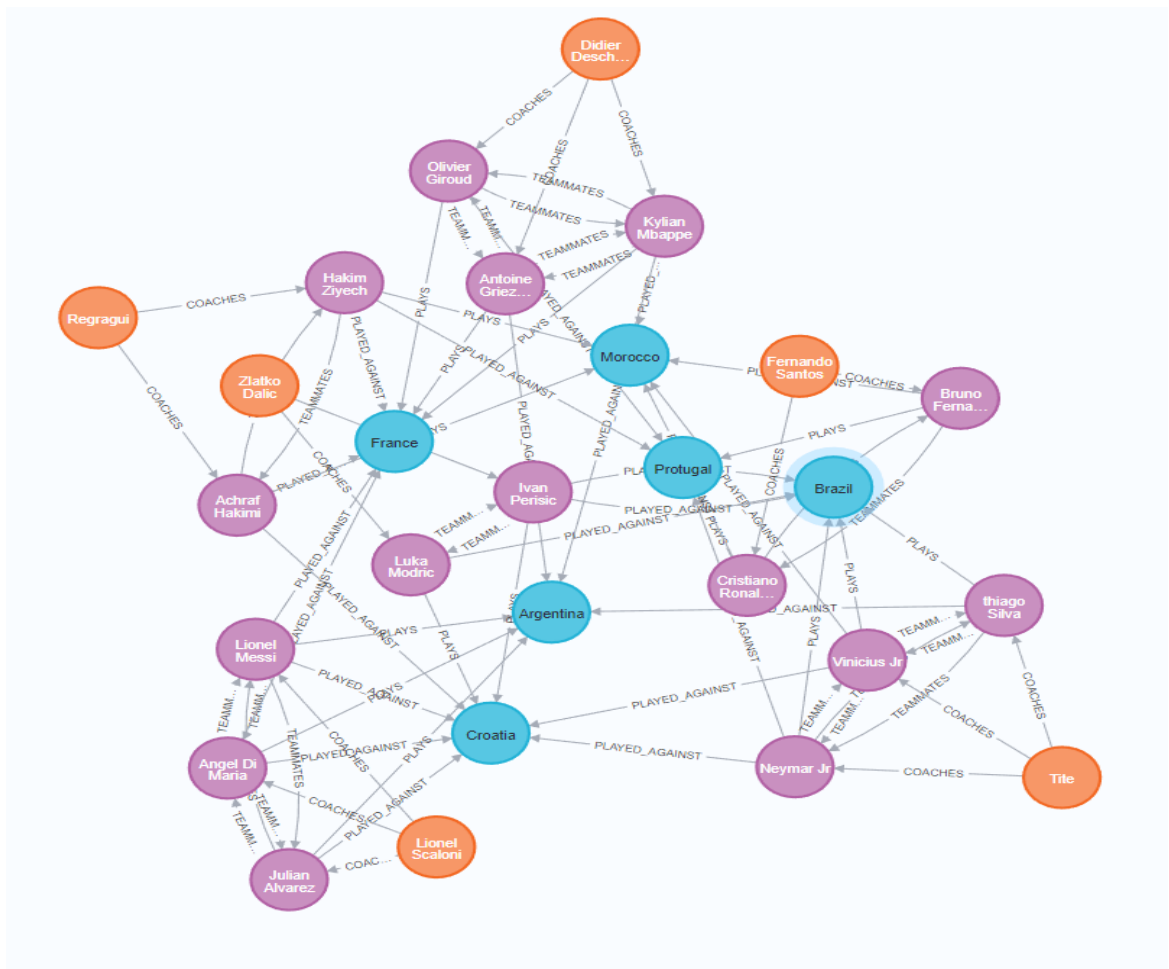
Node labels



Relationship types



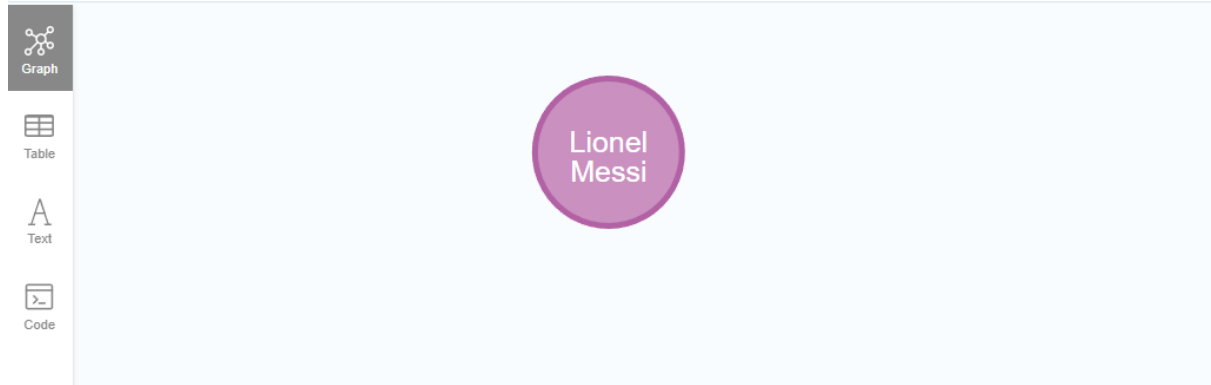
Displaying 27 nodes, 0 relationships.



Performing Different Queries

- In order to find the specification of a single node we can use this query.
- Here we have used a particular players name to find its node properties.

```
neo4j$ MATCH (player:PLAYER) WHERE player.name = "Lionel Messi" RETURN player
```



LIMIT:

- Limit is used to put constraint to the number of outputs to be viewed.
- To determine how many players might not be able to play future world cup whose age is greater than 35 and set limit as 5.

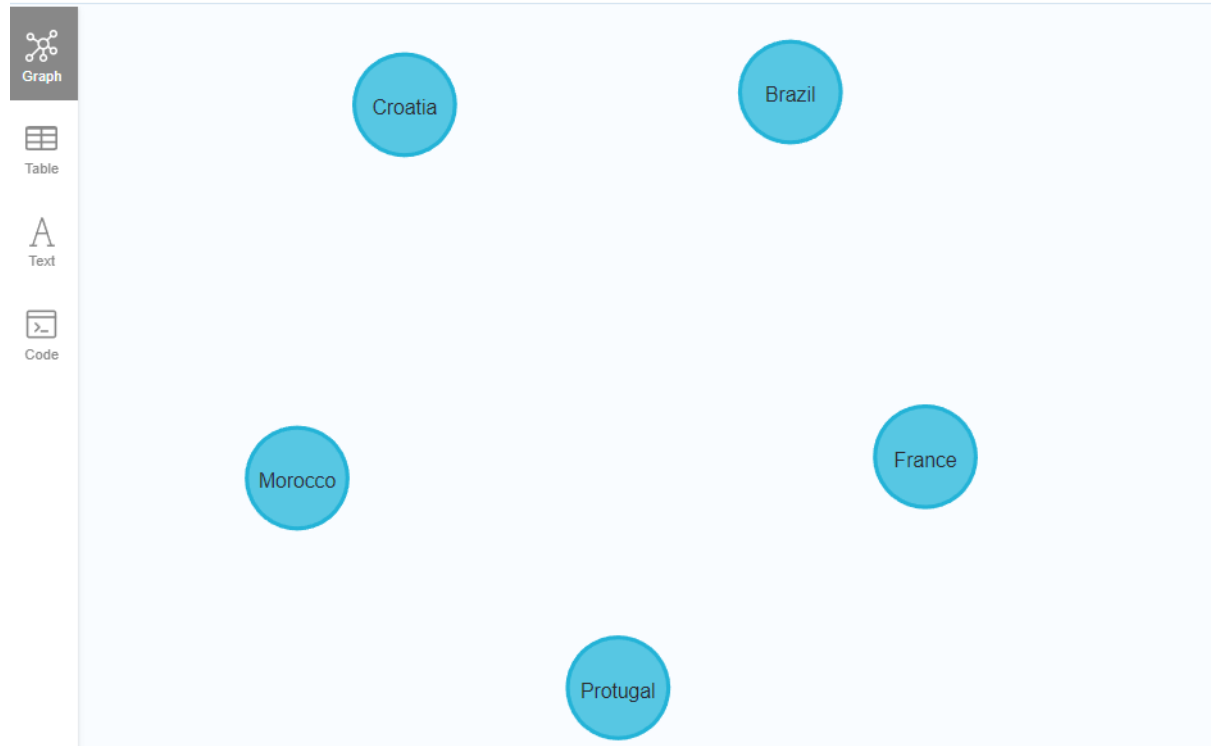
```
neo4j$ MATCH (player:PLAYER) WHERE player.age ≥ 35 RETURN player LIMIT 5
```



WHERE Clause:

- Where clause is used to specify any criteria to meet the users desired output.
- Here we have to find how many teams are in the dataset except Argentina.

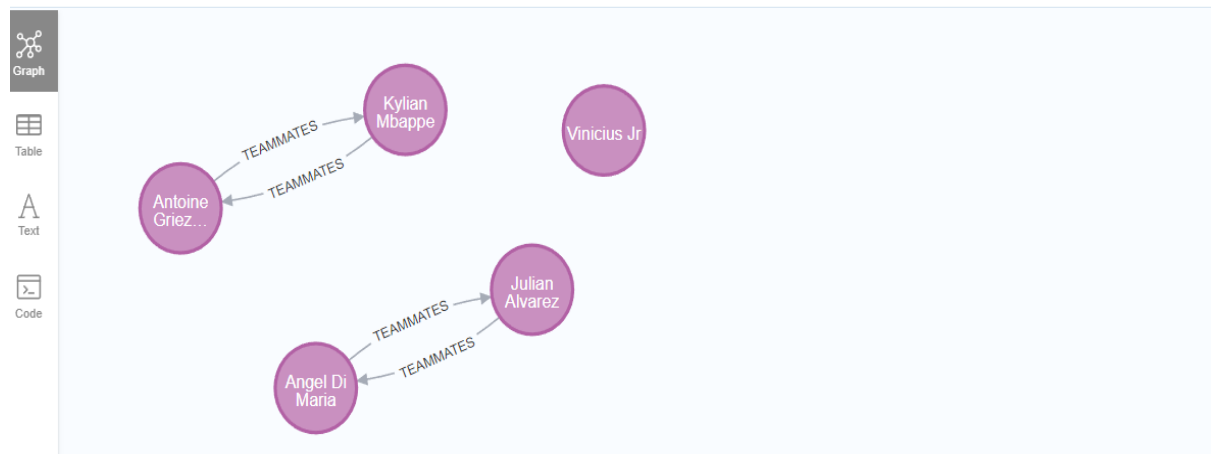
```
neo4j$ MATCH (team:TEAM) WHERE team.name <> "Argentina" RETURN team
```



AND Operator:

- AND operator is used to give multiple specific conditions in which all the conditions must be true to get the desired output.
- Here we have to determine players within constrained range of weight and height.

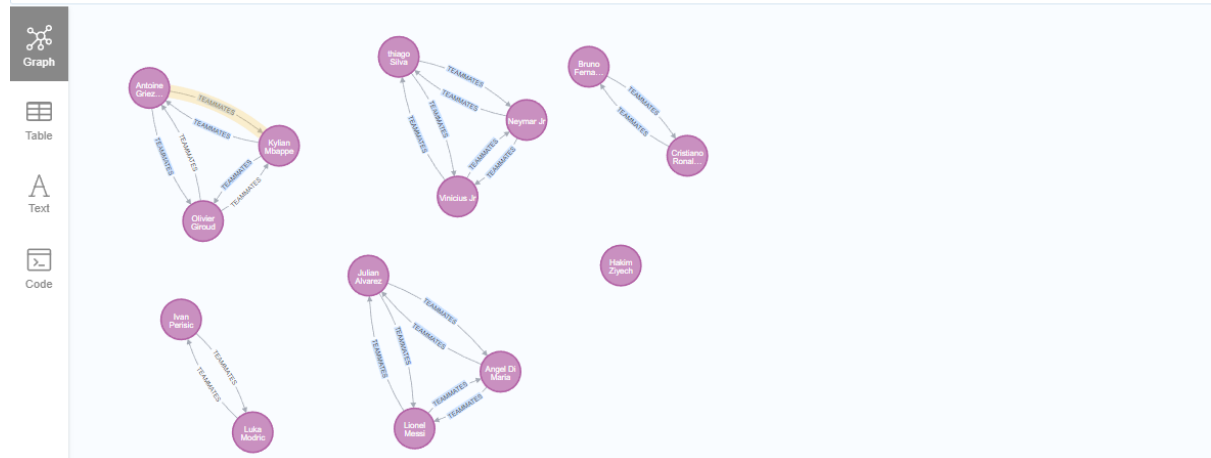
```
neo4j$ MATCH (player:PLAYER) WHERE player.weight ≥ 70 AND player.height ≤ 1.8 RETURN player
```



OR Operator:

- OR operator is used to give multiple specific conditions in which any of the conditions must be true to get the desired output.
- Here we have to determine players having weight more than 75 kg OR height less than 1.8m.

```
neo4j$ MATCH (player:PLAYER) WHERE player.weight ≥ 75 OR player.height ≤ 1.8 RETURN player
```



NOT Operator:

- NOT operator is used when we have to deny any specific condition.
- We have to determine young players from dataset which are NOT in the specified range.

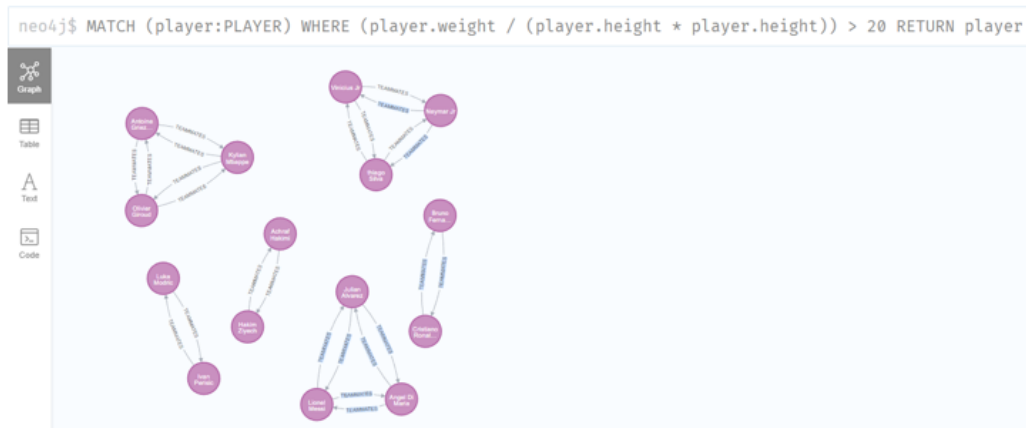
```
neo4j$ MATCH (player:PLAYER) WHERE NOT player.age ≥ 25 RETURN player
```



BMI Calculations

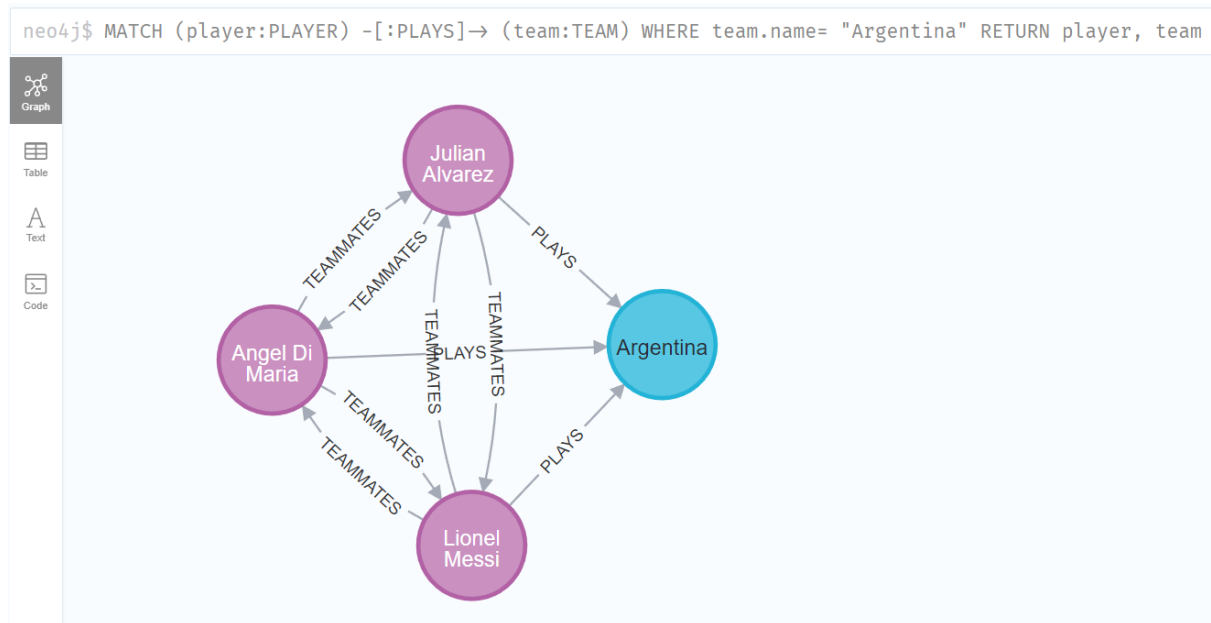
To find the fit player.

BMI: Calculated healthy players from our dataset on bmi parameter



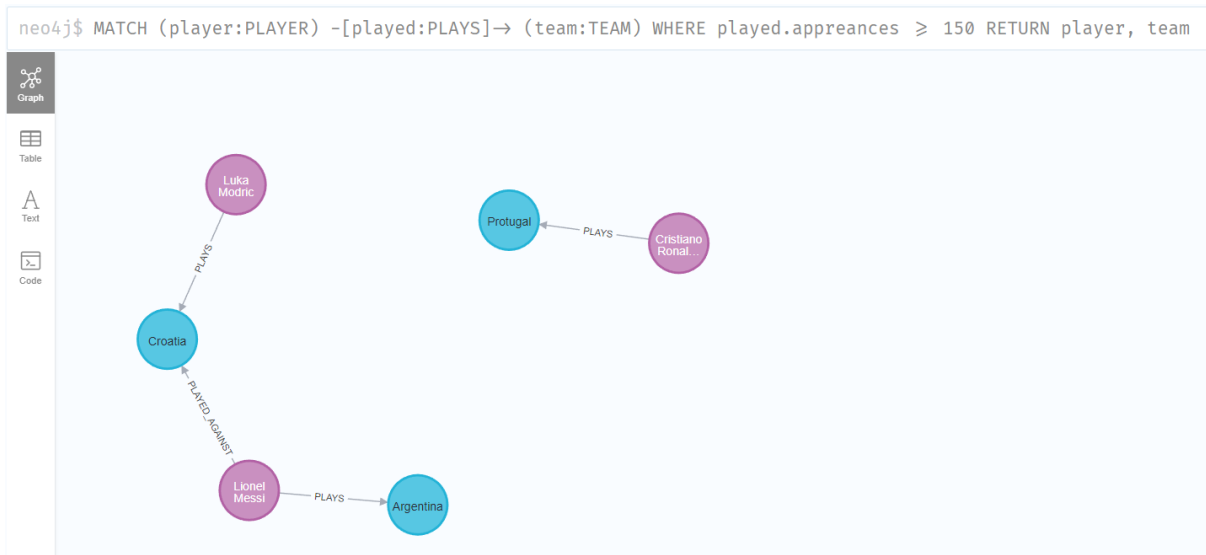
Query between relationship:

- Created query to find out the players which are playing for team 'Argentina'.
- There are three players from our dataset who play for Argentina team.



Query between relationship properties:

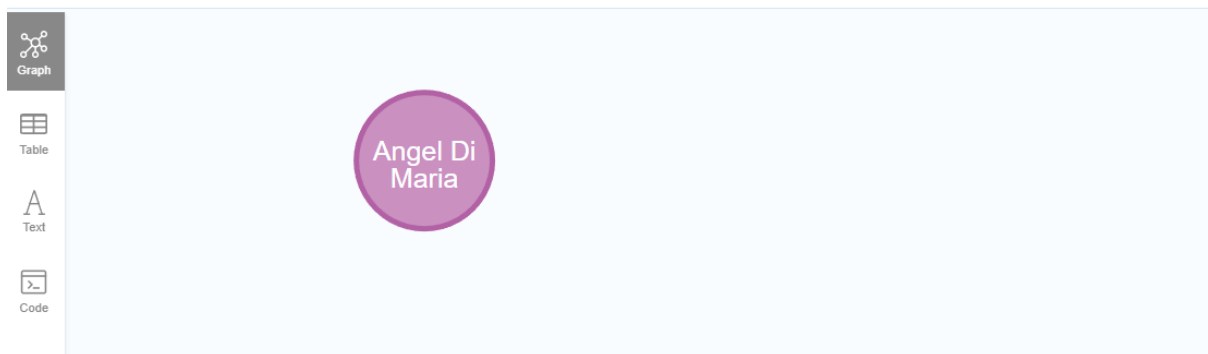
- To determine how many players have represented their country more than 150 times.
- Luka Modric, Lionel Messi and Cristiano Ronaldo have played more than 150 matches for their country.



Multiple Match Operator:

- Query is created with the help of multiple match in order to compare team mates on the basis of appearance as a property.
- We have determined which player played more than 100 times for their country which is team mate with Lionel Messi.

```
1 MATCH (player:PLAYER {name: "Lionel Messi"}) - [:TEAMMATES]→(teammate:PLAYER)
2 MATCH (teammate) - [played:PLAYS] → (team:TEAM)
3 WHERE played.appearances ≥ 100
4 RETURN teammate
```



Aggregate Functions (MIN):

- To determine minimum tackle committed by an individual player in single match.

<pre> 1 MATCH (player:PLAYER)- [matches:PLAYED_AGAINST]→ (:TEAM) 2 RETURN player.name, MIN(matches.tackle) AS Tackle 3 ORDER BY Tackle DESC </pre>		
Table	player.name	Tackle
1	"Luka Modric"	7
2	"Angel Di Maria"	5
3	"Lionel Messi"	4
4	"Kylian Mbappe"	3
5	"thiago Silva"	3
6	"Neymar Jr"	3
7		
Started streaming 15 records after 1 ms and completed after 2 ms.		

Aggregate Functions (MAX):

- To determine maximum goals scored by an individual player in single match.

<pre> 1 MATCH (player:PLAYER)- [matches:PLAYED_AGAINST]→ (:TEAM) 2 RETURN player.name, MAX(matches.goal) AS GOALS 3 ORDER BY GOALS DESC </pre>		
Table	player.name	GOALS
1	"Kylian Mbappe"	3
2	"Lionel Messi"	3
3	"Luka Modric"	3
4	"Olivier Giroud"	2
5	"Neymar Jr"	2
6	"Julian Alvarez"	2
7	"thiago Silva"	1
8	"Achraf Hakimi"	1
9	"Angel Di Maria"	1
Started streaming 15 records after 1 ms and completed after 2 ms.		

Aggregate Functions (SUM):

- To determine total goals scored by an individual player in all matches from dataset.

```
1 MATCH (player:PLAYER)- [matches:PLAYED_AGAINST]→ (:TEAM)
2 RETURN player.name, SUM(matches.goal) AS Total_Goals
3 ORDER BY Total_Goals DESC
```

	player.name	Total_Goals
1	"Kylian Mbappe"	4
2	"Neymar Jr"	4
3	"Lionel Messi"	4
4	"Luka Modric"	3
5	"Olivier Giroud"	2
6	"Angel Di Maria"	2
7		

Started streaming 15 records after 1 ms and completed after 2 ms.

Aggregate Functions (AVG):

- To determine average of total attempts made by an individual player per match.

```
1 MATCH (player:PLAYER)- [matches:PLAYED_AGAINST]→ (:TEAM)
2 RETURN player.name, AVG(matches.attempt)
```

	player.name	AVG(matches.attempt)
1	"Antoine Griezmann"	2.0
2	"Kylian Mbappe"	3.5
3	"thiago Silva"	4.0
4	"Olivier Giroud"	3.0
5	"Hakim Ziyech"	1.0
6	"Neymar Jr"	4.0
7	"Achraf Hakimi"	2.0
8	"Lionel Messi"	4.5
9	"Angel Di Maria"	1.5

Started streaming 15 records in less than 1 ms and completed after 1 ms.

Aggregate Functions (COUNT):

- To determine total matches played by an individual player from the dataset.

```
1 MATCH (player:PLAYER)- [matches:PLAYED_AGAINST]→ (:TEAM)
2 RETURN player.name, count(matches) As Total_Matches
3 ORDER BY Total_Matches DESC
4
```

	player.name	Total_Matches
1	"Kylian Mbappe"	2
2	"Hakim Ziyech"	2
3	"Neymar Jr"	2
4	"Achraf Hakimi"	2
5	"Lionel Messi"	2
6	"Angel Di Maria"	2
7		

Started streaming 15 records after 6 ms and completed after 6 ms.