Analysis of European Soccer League

Yun Xiao, Chia-Tsun Fan, Yeji Han, Sabrina Lee, Chu-Yueh Wang

Stage 1:

The organization we are choosing is the European Soccer League. The European Soccer League. It is an annual club football competition organized by the Union of European Football Associations (UEFA). European Soccer League is one of the most prestigious football tournaments in the world and the most prestigious club competition in European football.
The database we are using is a soccer tournament based on the EURO CUP 2016.
Below are the key conceptual elements of our data:

Entities and Definitions:

Key Entities	Definitions	Attributes	Examples
PLAYER	Player - an employee from a private entity which is a Football Club registered in the Football Association of the nation where he/she plays.	- player_id - team_id - jersey_no - player_name - posi_to_play - dt_of_bir - playing_club	E.g. Thomas Muller, the forward from Bayern
TEAM	Team - a group of 11 players who play for the country or club	- team_id - team_group - match_played - won - draw - lost - points	E.g. Bayern Munich
MATCH	Match - a game that was played, being played or about to be played by competing teams	- match_no - play_stage - play_date - results - referee_id - audence	E.g. The first game during Final
COACH	Coach (only means head coach) - a professional who instructs players on the skills they need to succeed in soccer.	- coach_id - team_id - coach_name	E.g. Joachim Low
REFEREE	Referee - a person who is responsible for the general supervision of the game and has the final authority on all rulings.	- referee_id - referee_name	E.g. Felix Brych

PLAYER_IN_OUT	Player in out - a player who competed in a match came into the field or went out from the field.	- match_no - player_id - in_out - time_in_out	E.g. I meaning a player came into the field

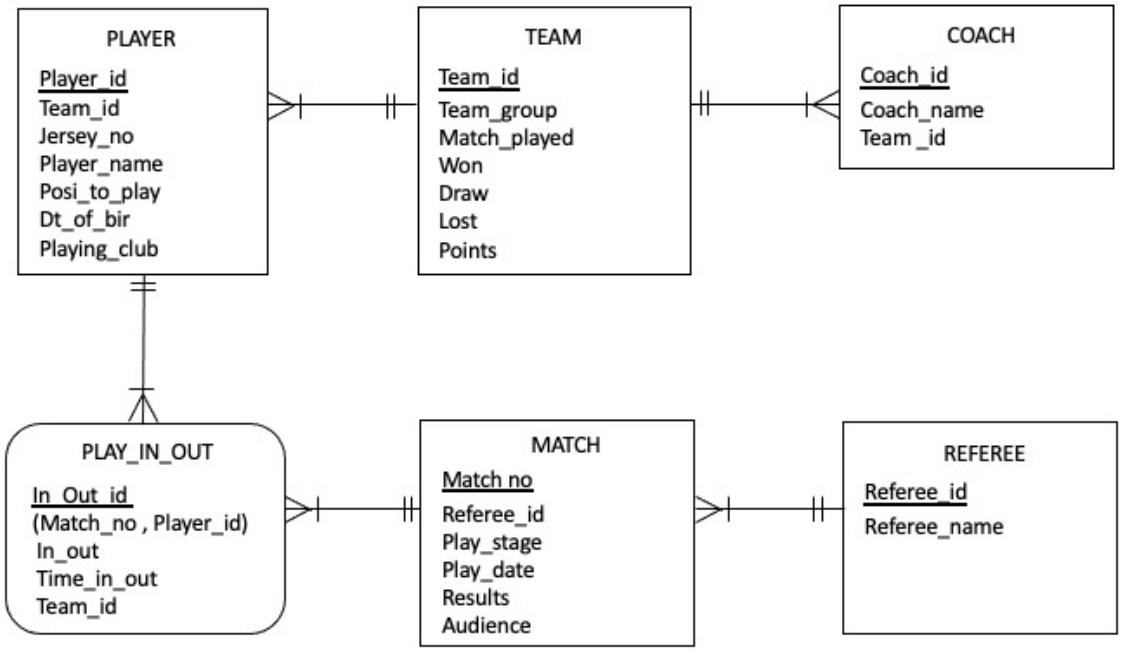
Relationships:

PLAYER-TEAM	A TEAM has one or many PLAYERs. A PLAYER belongs to one and only one TEAM.
PLAYER-COACH	A COACH leads one or many PLAYERs. A PLAYER is led by one or many COACHes.
TEAM-COACH	A TEAM is led by one or more COACH. A COACH leads one and only one TEAM.
MATCH-REFEREE	A MATCH has one and only one head REFEREE. A head REFEREE is in charge of one or many MATCH.
PLAYER-MATCH	A PLAYER attends one or many MATCHes. A MATCH is held by one or many PLAYERs.

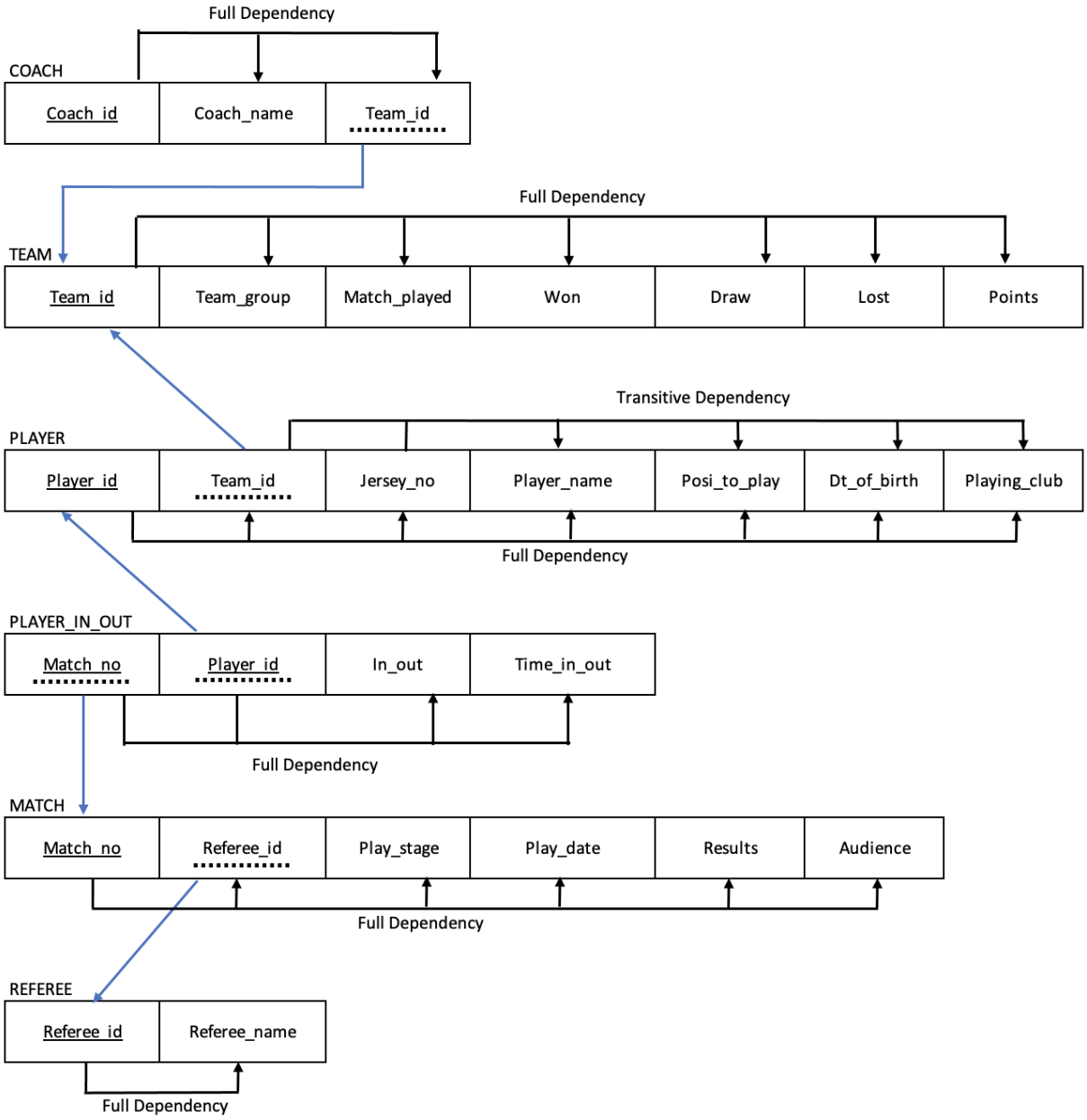
Stage 2:

> Draw the relational schema based on your ERD, or if you have actual data tables, draw the relational schema based on your actual data tables.

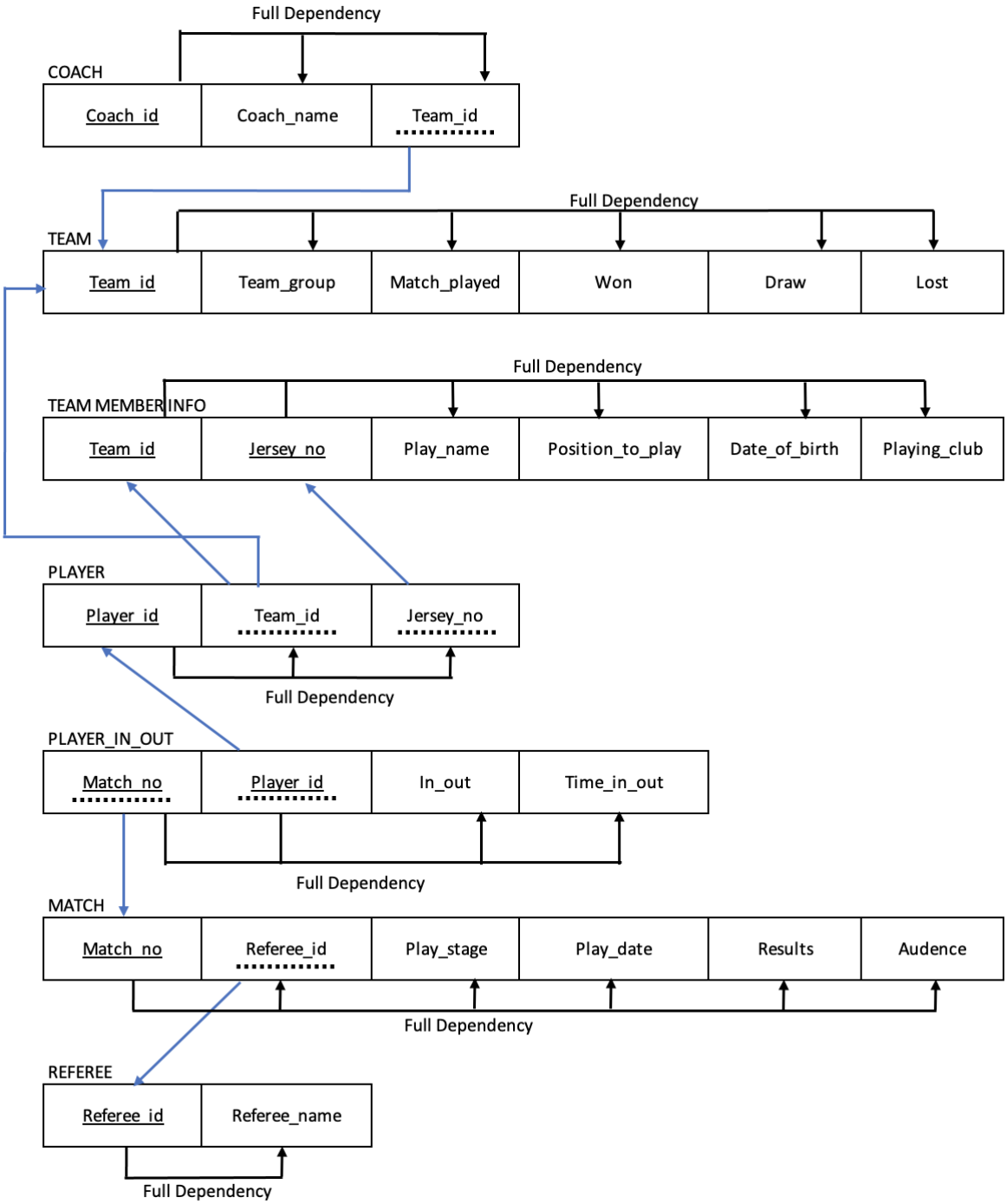
ERD:



> Draw the as-is dependency diagram and identify types of dependencies.



> Do you have tables in 3rd normal form? If not, convert your tables to 3rd normal form.



Stage 3:
> Screenshot of your queries and the results
Part 1 Number of coaches and teams' performance (query 1 - 3)
1.
Firstly, we analyzed if every team hired only one coach. From the result, nearly all the teams only have one coach except for team 1210. It has 2 coaches, which is rare for soccer.
Query:

```
select t.team_id, count(coach_id) as cnt
from team t
left join coach c
on(t.team_id=c.team_id)
group by t.team_id
having cnt>1;
```

Result:

team_id	cnt
1210	2

The importance of each query for managerial decision making:
The query counts the number of coaches for different teams and selects those with 2 or more than 2 coaches (team 1210), which lays the foundation for further analytics.

2.
Then we used queries to find the two coaches are Lars Lagerback and Heimir Hallgrimson.
Query:

```
select team_id,coach_name
from coach
where team_id in (select t.team_id
from team t
left join coach c
on(t.team_id=c.team_id)
group by t.team_id
having count(coach_id)>1);
```

Result:

team_id	coach_name
1210	Lars Lagerback
1210	Heimir Hallgrimsson

The team is Iceland, which was not outstanding before EURO CUP 2016. We'd like to see how the team performed under the co-coach system in 2016.

The importance of each query for managerial decision making:
By joining two tables and selecting the names of coaches, the query lets us know who are the two coaches of team 1210. Because in EURO CUP 2016, only team 1210 (Iceland) had a co-coach system, the query leads us to a further analysis of how the system did for Iceland in the game.

3.
In play_stage, G for Group stage, R for Round of 16 stage, Q for Quarter final stage, S for Semi Final stage, and F for Final. Hence, Iceland did well in EURO CUP 2016. It reached the quarterfinals.

Query:

```
select match_no, play_stage, results, play_date
from match_
where match_no in(select match_no
from player_in_out
where team_id=1210);
```

Result:

match_no	play_stage	results	play_date
12	G	DRAW	6/15/16
23	G	DRAW	6/18/16
33	G	WIN	6/22/16
44	R	WIN	6/28/16
48	Q	WIN	7/4/16

The importance of each query for managerial decision making:
In this part, the query selects play_stage and results of Iceland (team 1210). Because Iceland didn't perform well in soccer, this time it entered quarterfinals which implies that the co-coach system may have some positive impact on the improvement of the team. Therefore, step by step, the queries of #1, #2 and #3 let us know the unique co-coach system of Iceland, EURO CUP 2016's dark horse, may have some positive influence on its success. The insight is able to lead to further analysis for managerial decision making such as if co-coach is superior to the common single coach system.

Part 2 The age of teams (query 4)

4.

Query:

This query lets us know how many years the team has been around. In order to calculate the age of each team, we utilize the column “date_of_birth” to calculate.

```
select team_id, round(avg(year(current_date()) - year(dt_of_bir)),0) as avg_year
from player
group by team_id
order by avg_year desc;
```

Results and Findings:

We found that each team is about 30-36 years old. The oldest team in this dataset is team 1215, which has been established for 36 years.

team_id	avg_year
1215	36
1205	35
1211	35
1212	35
1216	35
1217	35
1218	35
1201	34
1202	34
1207	34
1209	34
1210	34
1213	34
1214	34
1219	34
1204	33
1203	32
1206	32
1208	32

The importance of each query for managerial decision making:

This query lets us know how many years the team has been around and learn about the characteristics of the team. For example, if the team is young, they may receive high popularity. On the other hand, if the team has been established for a long time, they may have a higher winning rate due to their great experience. Also, the organization can compare the performance of teams of different ages to see if younger or older teams perform better and adjust strategies accordingly.

Part 3 The times of teams winning and teams' popularity (query 5)
5.
Query:
This query shows how many times the team has won and how popular it is. In order to know the number of audiences and the number of wins of each team, here we join the two tables: "match" and the association table "player_in_out".

```
# The team with the most wins and the most popular team
select team_id, count(results) as Win,
round(avg(audence),2) as AverageAudience
from match_
left join player_in_out
on match_.match_no = player_in_out.match_no
where results = 'win'
group by team_id
order by win desc;
```

Results and Findings:
As we can see, team 1207 has the most wins and has the highest audience. Therefore, team 1207 is the relatively popular and strongest team out there.

team_id	Win	AverageAudience
1207	36	68640.17
1224	32	41350.50
1211	30	48841.00
1203	30	40753.80
1208	26	47780.23

The importance of each query for managerial decision making:
This query could help the organization to know how many times the team has won and how popular it is. For the team with relatively large audiences, the organization should focus on the management of crowds at sporting events and maintain the food quality of the concession stand to support such a large crowd. As for the team that has won many times but has a lower audience, the organization should find out the reason why the team is not that popular so that they can develop marketing strategies to raise the popularity of that team.

> Scope for extension. If you had more time and resources, how would you extend the current project?

We will search for more data about different coach systems and the results of more soccer games. And then we are going to analyze the relationship between coach systems and the team's performance.

Reference:

https://www.w3resource.com/sql-exercises/soccer-database-exercise/index.php