COMP 3005 Project Report

By: Alan Zhou, Mengxuan Shi, and Ethan Bauer

1 Conceptual Design

Following the Statsbomb documentation, we stored the events data by event type, with a main event table to store the common attributes of the events.

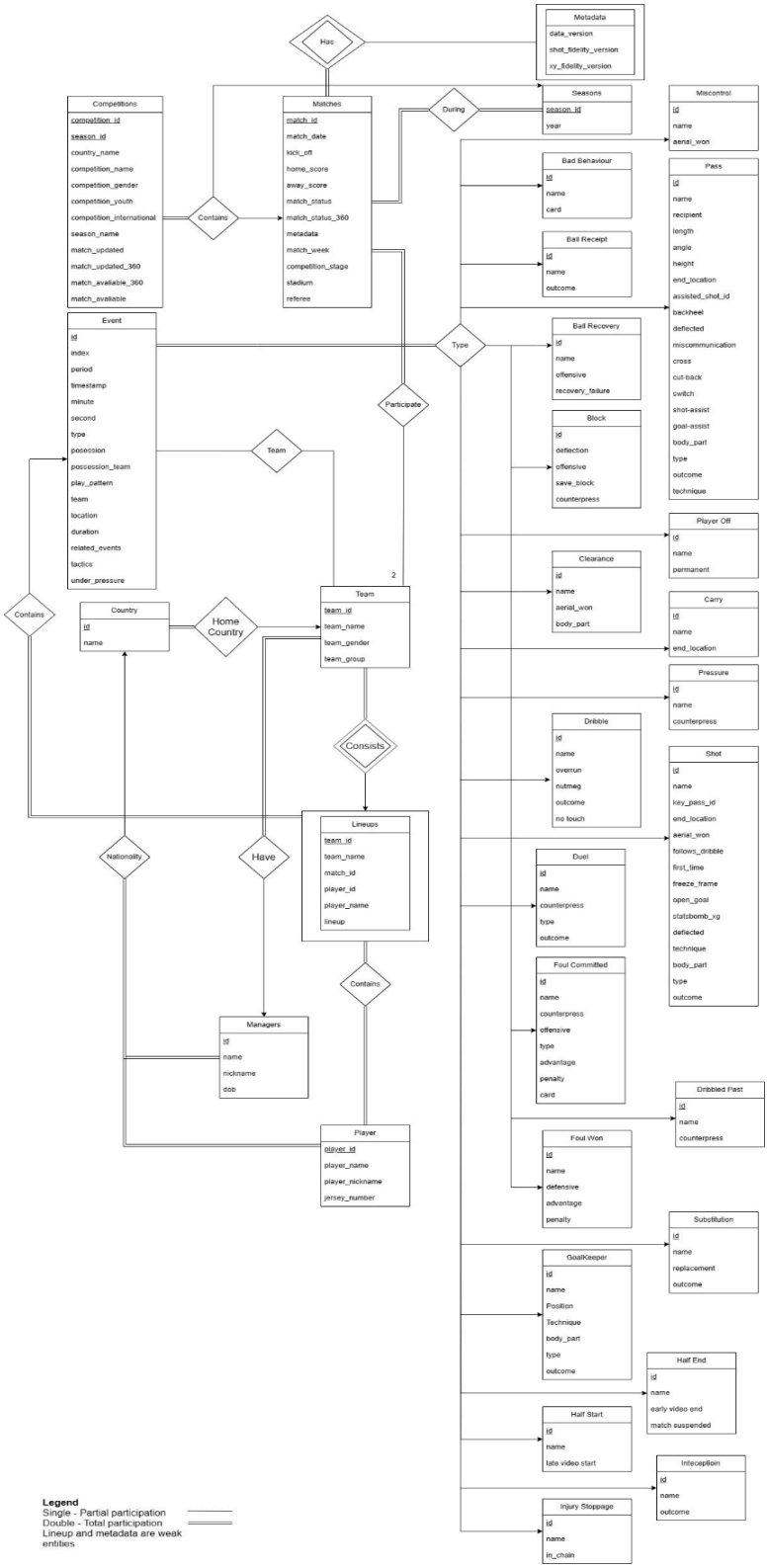
The data from \open-data\data\matches is parsed into the competition table, allowing querying attributes related to the matches

We developed a Python program to filter the events related to both La Liga and Premier League in the 2003/2004, 2020/2021, 2019/2020, and 2018/2019 seasons. This significantly reduces the workload of parsing data parsing and the resulting database size.

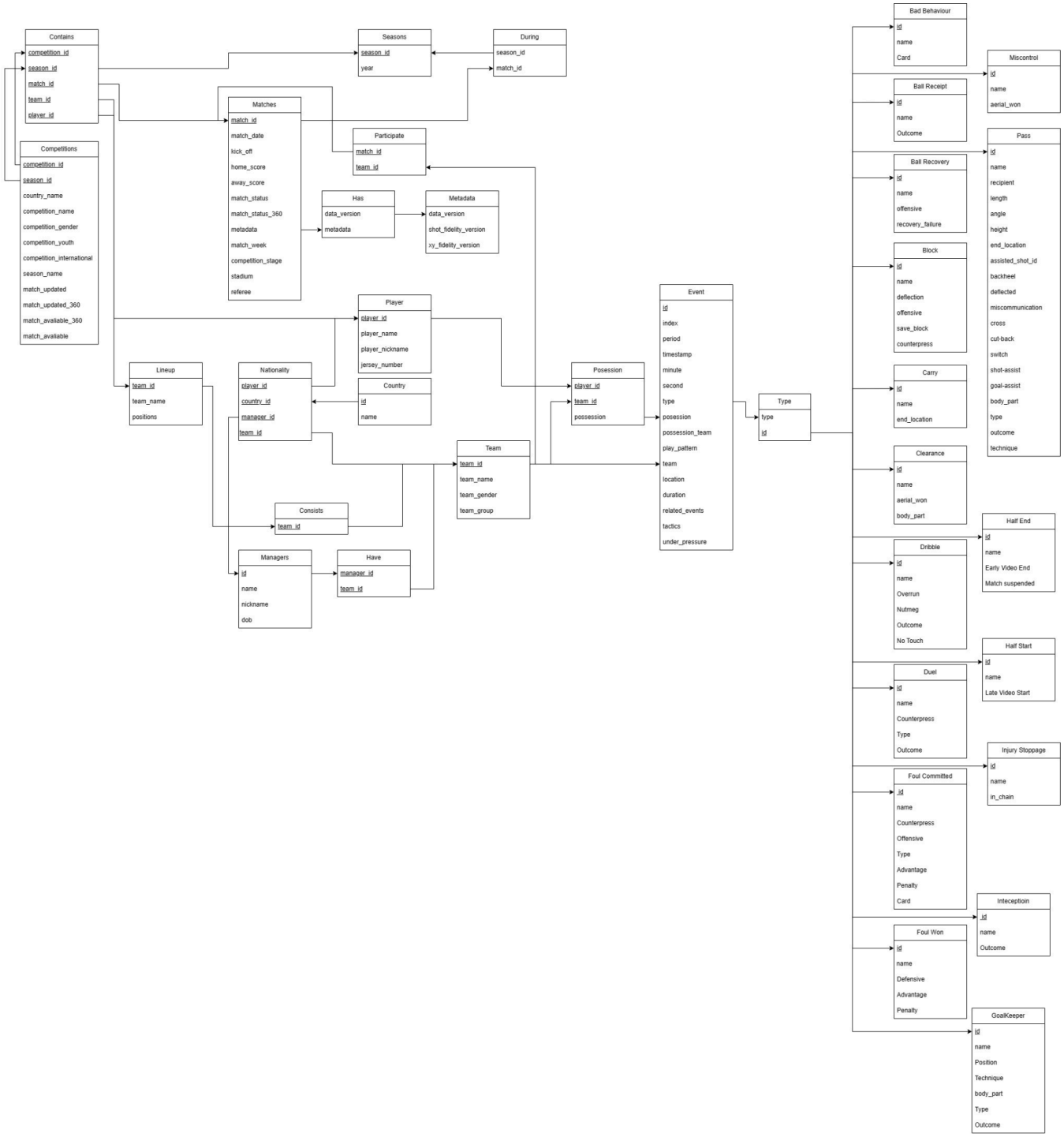
Events within a match are obtained from the directory \open-data\data\events, where each event file corresponds to the match id. All events and matches are uniquely identifiable by its id from the Statsbomb dataset.

We assumed the matches participated in the team and the lineups contain the players.

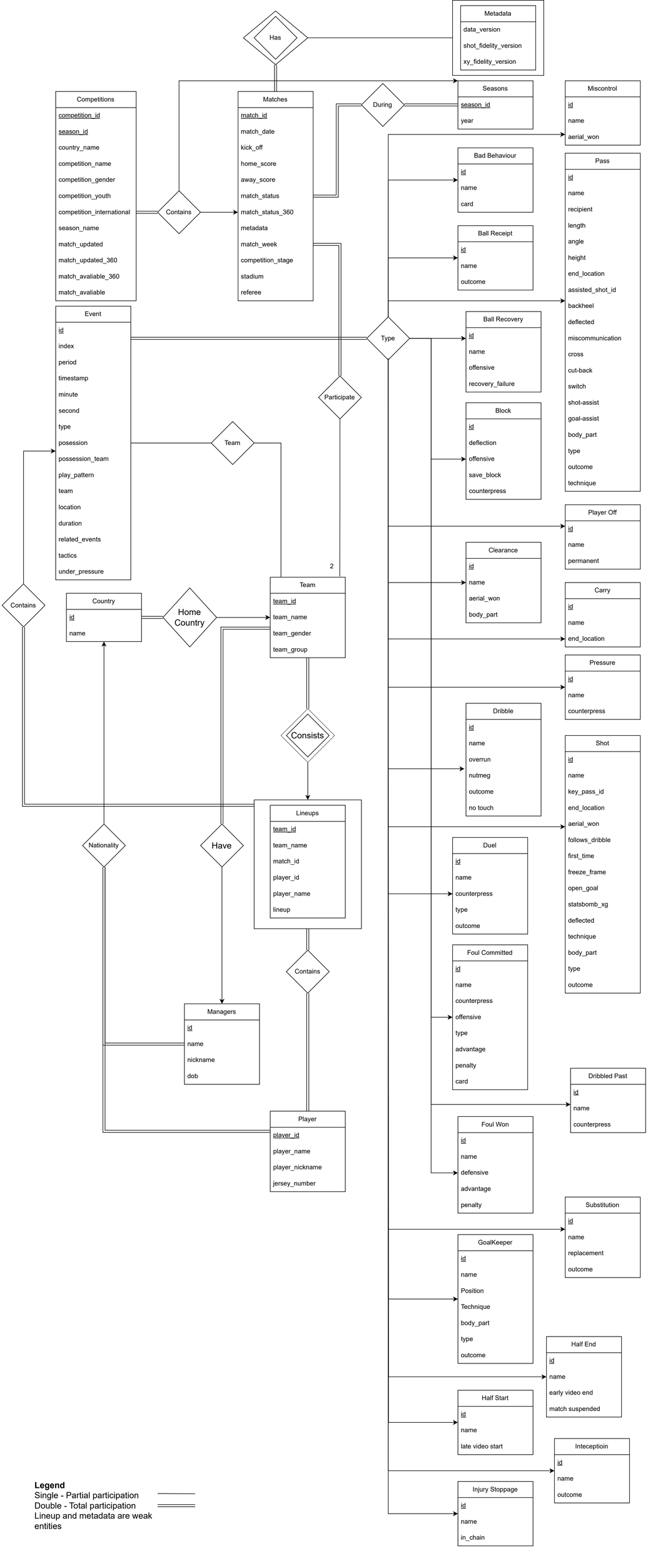
2 Reduction to Relation Schema

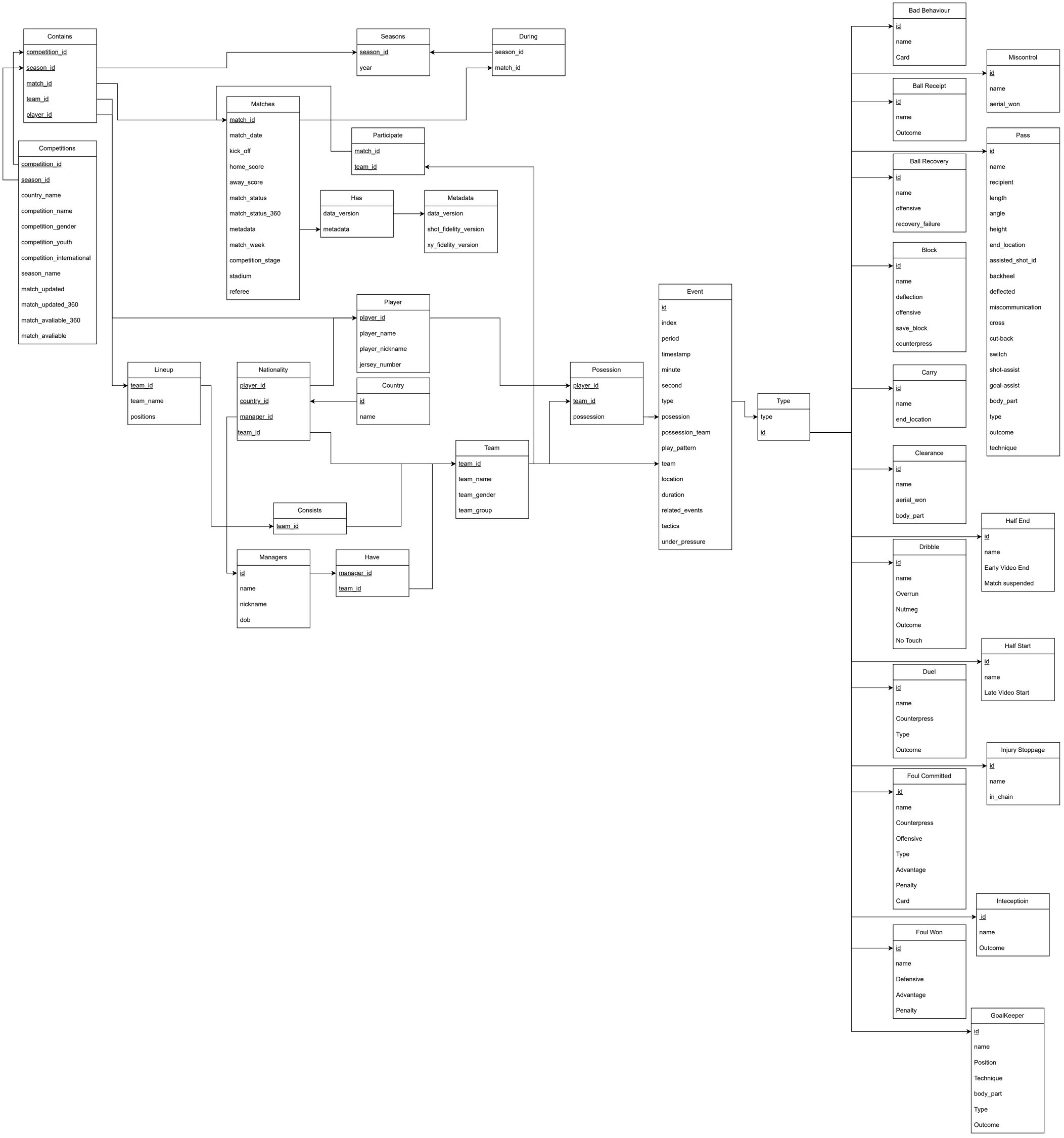


3 Database Schema Diagram



[URL to diagrams]


Has
Metadata
data_version
shot_fidelity_version
xy_fidelity_version
Competitions
competition_id
season_id
country_name
competition_name
competition_gender
competition_youth
competition_international
season_name
match_updated
match_updated_360
match_avaliable_360
match_avaliable
Contains
Matches
match_id
match_date
kick_off
home_score
away_score
match_status
match_status_360
metadata
match_week
competition_stage
stadium
referee
During
Seasons
season_id
year
Miscontrol
id
name
aerial_won
Bad Behaviour
id
name
card
Ball Receipt
id
name
outcome
Pass
id
name
recipient
length
angle
height
end_location
assisted_shot_id
backheel
deflected
miscommunication
cross
cut-back
switch
shot-assist
goal-assist
body_part
type
outcome
technique
Event
id
index
period
timestamp
minute
second
type
posession
possession_team
play_pattern
team
location
duration
related_events
tactics
under_pressure
Type
Ball Recovery
id
name
offensive
recovery_failure
Participate
Block
id
deflection
offensive
save_block
counterpress
Team
Player Off
id
name
permanent
Clearance
id
name
aerial_won
body_part
2
Contains
Country
id
name
Home Country
Team
team_id
team_name
team_gender
team_group
Carry
id
name
end_location
Pressure
id
name
counterpress
Consists
Dribble
id
name
overrun
nutmeg
outcome
no touch
Shot
id
name
key_pass_id
end_location
aerial_won
follows_dribble
first_time
freeze_frame
open_goal
statsbomb_xg
deflected
technique
body_part
type
outcome
Lineups
team_id
team_name
match_id
player_id
player_name
lineup
Nationality
Have
Duel
id
name
counterpress
type
outcome
Foul Committed
id
name
counterpress
offensive
type
advantage
penalty
card
Contains
Managers
id
name
nickname
dob
Dribbled Past
id
name
counterpress
Player
player_id
player_name
player_nickname
jersey_number
Foul Won
id
name
defensive
advantage
penalty
Substitution
id
name
replacement
outcome
GoalKeeper
id
name
Position
Technique
body_part
type
outcome
Half End
id
name
early video end
match suspended
Half Start
id
name
late video start
Inteceptioin
id
name
outcome
Injury Stoppage
id
name
in_chain
Legend
Single - Partial participation
Double - Total participation
Lineup and metadata are weak entities


Contains
competition_id
season_id
match_id
team_id
player_id
Seasons
season_id
year
During
season_id
match_id
Matches
match_id
match_date
kick_off
home_score
away_score
match_status
match_status_360
metadata
match_week
competition_stage
stadium
referee
Participate
match_id
team_id
Competitions
competition_id
season_id
country_name
competition_name
competition_gender
competition_youth
competition_international
season_name
match_updated
match_updated_360
match_avaliable_360
match_avaliable
Has
data_version
metadata
Metadata
data_version
shot_fidelity_version
xy_fidelity_version
Player
player_id
player_name
player_nickname
jersey_number
Lineup
team_id
team_name
positions
Nationality
player_id
country_id
manager_id
team_id
Country
id
name
Team
team_id
team_name
team_gender
team_group
Consists
team_id
Managers
id
name
nickname
dob
Have
manager_id
team_id
Posession
player_id
team_id
possession
Event
id
index
period
timestamp
minute
second
type
posession
possession_team
play_pattern
team
location
duration
related_events
tactics
under_pressure
Type
type
id
Bad Behaviour
id
name
Card
Miscontrol
id
name
aerial_won
Ball Receipt
id
name
Outcome
Pass
id
name
recipient
length
angle
height
end_location
assisted_shot_id
backheel
deflected
miscommunication
cross
cut-back
switch
shot-assist
goal-assist
body_part
type
outcome
technique
Ball Recovery
id
name
offensive
recovery_failure
Block
id
name
deflection
offensive
save_block
counterpress
Carry
id
name
end_location
Clearance
id
name
aerial_won
body_part
Half End
id
name
Early Video End
Match suspended
Dribble
id
name
Overrun
Nutmeg
Outcome
No Touch
Half Start
id
name
Late Video Start
Duel
id
name
Counterpress
Type
Outcome
Injury Stoppage
id
name
in_chain
Foul Committed
id
name
Counterpress
Offensive
Type
Advantage
Penalty
Card
Inteceptioin
id
name
Outcome
Foul Won
id
name
Defensive
Advantage
Penalty
GoalKeeper
id
name
Position
Technique
body_part
Type
Outcome