



## PROJECT BIG DATA

**Shan Hu**

Studentnummer: 500684659

**Klas:**

FDMCI\_IVBD1-6

**Michael Visser**

Studentnummer: 500823669

**Docent:**

Peter Odenhoven

**Wang Di Jiang**

Studentnummer: 500826958

**Abdellah Amnad**

Studentnummer: 500799703

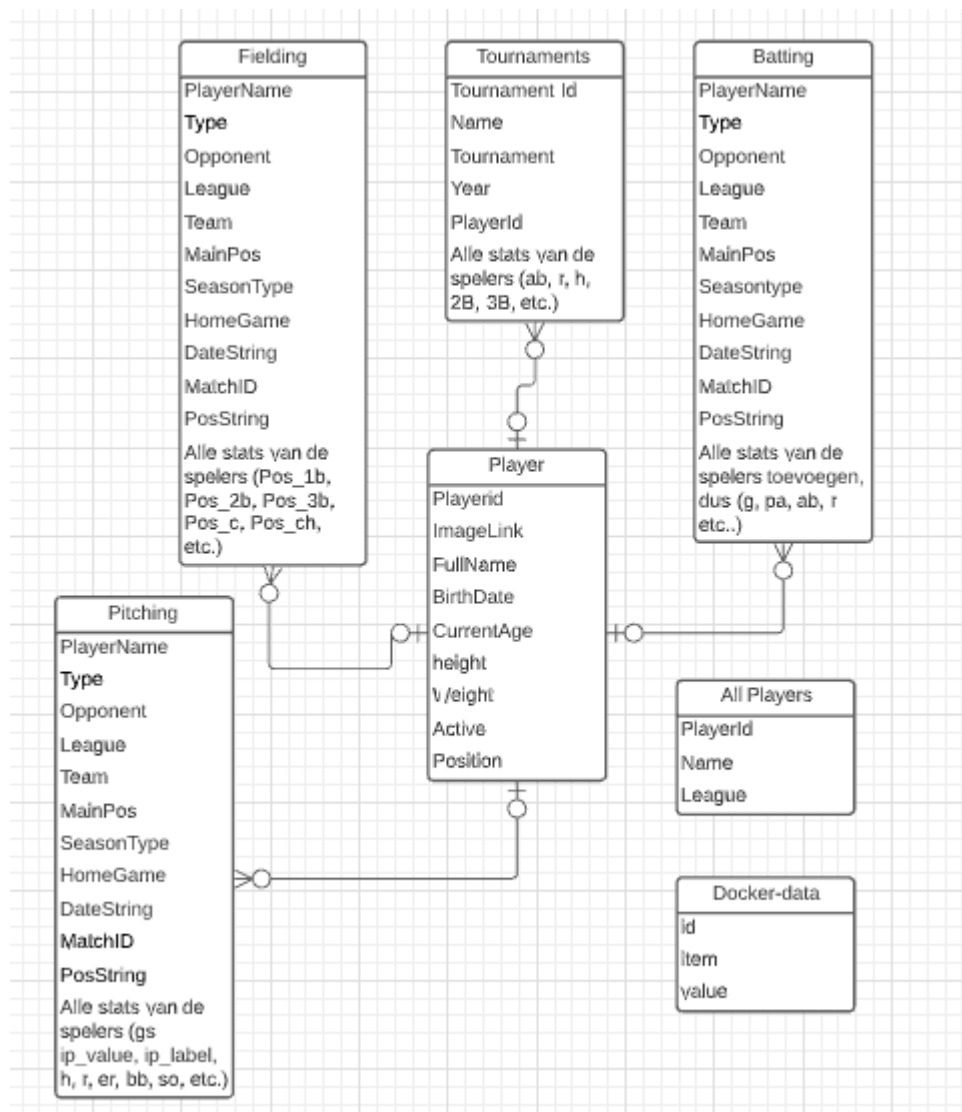
**Semih Simsir**

Studentnummer: 500830526

## Inleiding

Na overleg met Peter is naar voren gekomen dat de huidige database toe is aan vernieuwing. Momenteel is er veel irrelevante en dubbele data aanwezig die niet worden gebruikt. Daarnaast is gebleken dat de voorgaande database vanuit een mongoDB (niet relationele database) is omgezet naar een MySQL database waardoor er geen relaties aanwezig zijn in de database. Het is lastig om nieuwe data in de database toe te voegen, omdat er geen relaties tussen de tabellen aanwezig zijn. Op verzoek van Peter hebben wij een potentiële nieuw database ontwerp gemaakt, om het volgende team op weg te helpen.

## Database ERD



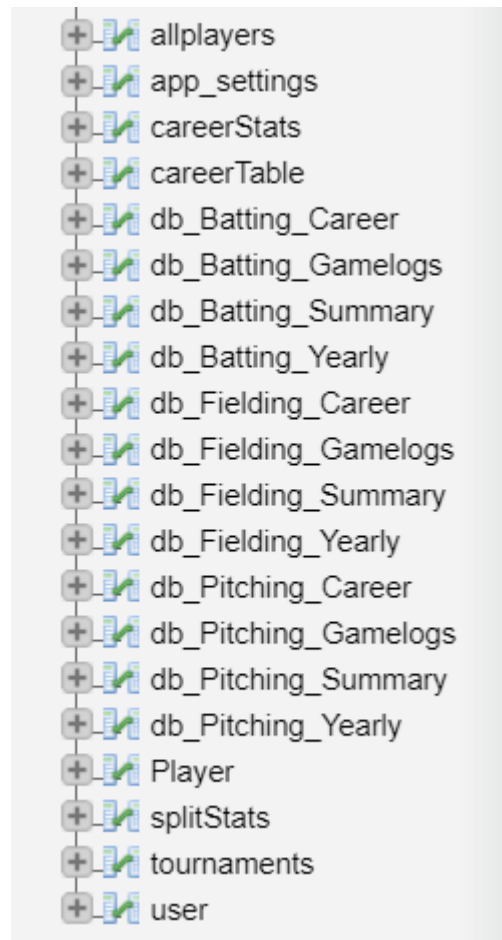
Figuur 1

## Toelichting

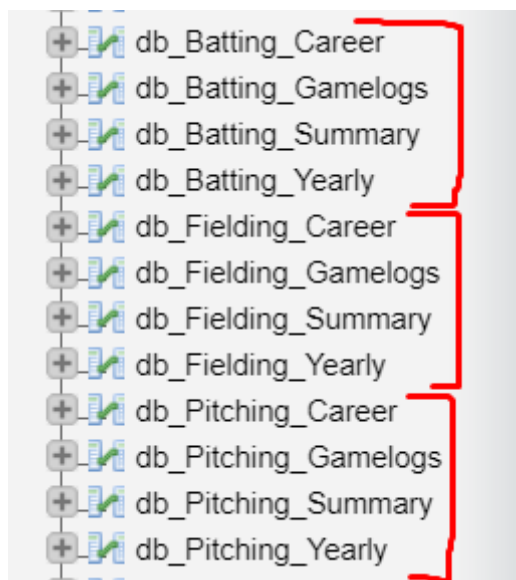
In figuur 1.1 is de huidige database te zien. We hebben ervoor gekozen om alle db\_Batting, db\_Fielding en db\_Pitching (zie figuur 1.2) samen te voegen in drie verschillende tabellen. Dit is te zien in figuur 1 onder de tabellen Fielding, Pitching en Batting. Daarbij zijn alle attributen van de huidige database in overgenomen en een nieuwe kolom "Type" toegevoegd om te bepalen of het gaat om een Career, Gamelog, Summary of Yearly. Verder hebben we ervoor gekozen tussen de tabellen one to many relaties toe te passen waarbij het optioneel is of een speler deel uitmaakt van batting, fielding, pitching of tournaments. Zo heeft bijvoorbeeld een speler met alleen battingdata, geen data voor andere tabellen zoals fielding of pitching.

De All Players tabel staat los van alle andere tabellen, omdat deze tabel geen statistieken bevat van de spelers, maar enkel de spelersnaam, id en league. Deze spelers worden nog niet gebruikt in de applicatie, pas als ze via de add player page worden toegevoegd. Vandaar dat zij geen relatie hebben met de andere tabellen.

De Docker-data tabel staat ook los van alle andere tabellen. Deze tabel wordt alleen gebruikt om in te zien wanneer docker heeft gerund.



Figuur 1.1



Figuur 1.2