Data Management and Database Design

Physical Data Model and Social Media

Quan Wen

NBA_Player stats
Team Names
Players Names
Points Made
Other stats

Producer: Players
Players Names
Players Tweets
Players ID
retweets
Likes

Company:Team
Team Names
TeamTweets
Team ID
retweets
Likes

Consumer: #Players
Fans Names
Fans Tweets#PlayersNames
Fans ID
retweets
Likes
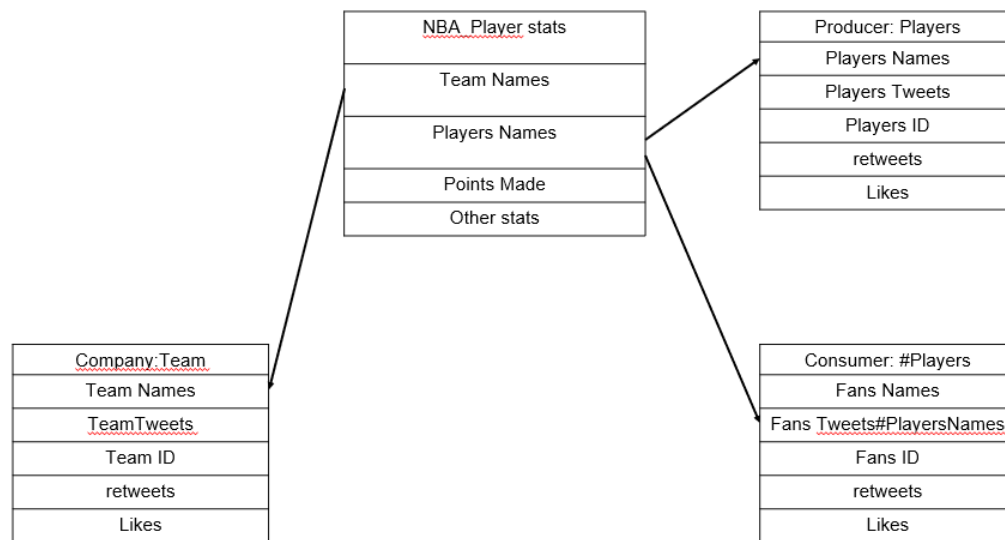
1.1 Purpose

I need to use API method to extract data from Twitter and other website and analyze data. For example, find out which team has most likes so that I can know which team is more famous in Twitter, find out what posts the players post in 24 hours.

1.2 What are the ranges, data types and format of all of the attributes in your entities?

To extract the newest information of the NBA players, I choose twitter as the resource of data.

For Producer, I extract NBA Players on Twitter, his name, the number of retweets of his tweet, His newest tweets in 24 hours , the number of likes of players tweets.

For Company, I extract NBA Teams on Twitter , Teams name, Team Tweets, Team ID and Team retweets, the number of likes of teams tweets.

For consumers, I extract fans tweets whose tweets include #Players Name on Twitter. On twitter I extract Fans Names, Fans ID , the number of retweets of their posts and likes.

1.3 When to use an entity versus attribute?

For Domain, according to the datas , I can find out that in the 2019 season ,how many two points have been made, his name, his team and other stats that we can know how the performance of the players.

For company , we can find out how the NBA Team post about themselves on Twitter. According to their posts we can know that how public respond to them by viewing the number of likes in Twitter.

For Producer, we can find out how the NBA Players post about themselves on Twitter. According to their posts we can know that how public respond to them by viewing the number of likes in Twitter.

For Consumers, we can find out what kind of topics they are talking about and find out which players they pay most attention to according to the number of retweets of the posts.

1.4 When should you use an entity or relationship, and placement of attributes? (Example: a manager could be modeled as either)?

Among these three entities(company , producer, fans), relationship is clear. When we talk about the players, we want to know how public evaluate this player , meanwhile , we also want to find to which company does the player belong. So relationship is quite clear.
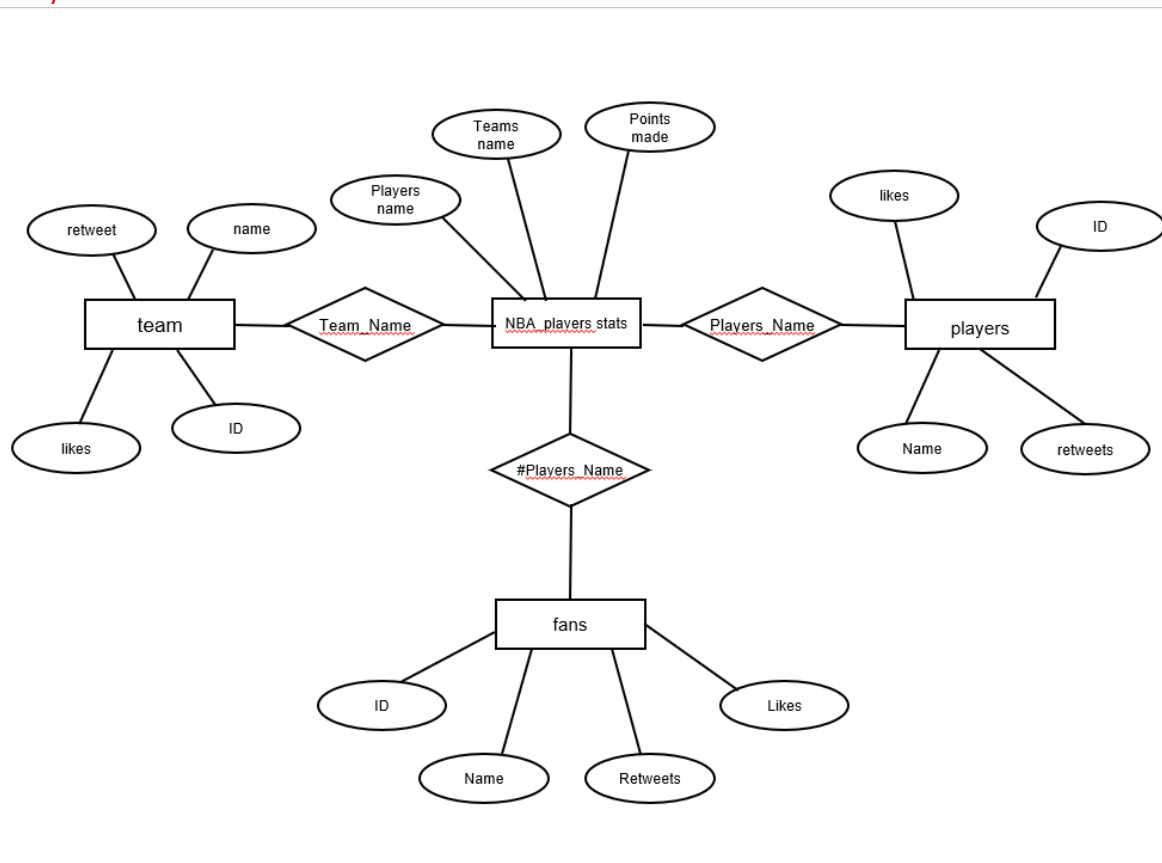
1.5 How did you choose your keys? Which are unique?

Keys: Team_name , Players name .

Because twitter give all of the Account , so I use thses names as the primary key and they are unique.

Part II SQLPart

Physical Model

2.1. What user posted this (e.g. tweet, facebook post, IG post, etc.)?
In query I transferred two results, one is Players' User_Name, the other is id from the Producer database, you can see which player sent posts.

2.2. When did the user post this (e.g. tweet, facebook post, IG post, etc.)?
In this query, I extract the date of tweets based on the User_Name in order to find out when did the user post this.

2.3. What posts has this user posted in the past 24 hours?
I use the select where command to extract the data and the User_Name and posts to find out which user and there posts in the past 24 hours

2.4. How many post has this user posted in the past 24 hours?
I use the Count AS num command to extract a particular user name and find out how many tweets this user posted in the past 24 hours.

2.5. What keywords/ hashtags are popular?
I first combine the Consumer and DomainToConsumer database together and then use the Max command to determine what hashtag is the most popular one.

2.6 What posts are popular?
In query I extract id, retweets, posts from Producer database where the number of retweets is higher than 1000 in order to find out what posts are popular.