Sailesh Sikdar

NBA Records Database

Chris Brogly

CS 338

University of Waterloo

Database Overview

I’m someone who is a basketball fan, playing when I have time and keeping up with the NBA. When I was given this project, I soon realized that it would be monotonous, and that it’s best to choose a topic of interest. I decided I’d make a table regarding NBA records, as there is a market for the records and lists of this type. A website known as Basketball Reference has statistics and logs of game box scores, teams and their results, players, their results and more. It is often used by the NBA community when arguing amongst each other to pass the time. Databases with advanced stats are in high demand as the NBA community as a whole and NBA teams are focusing on more of a data analytics approach to the game – hence why the Houston Rockets take so many three pointers a game. However, as advanced analytics are beyond the scope of my knowledge, they are not in this database.

The database stores general NBA records the last 10 years. So for each season, you have a Western Conference standings table as well as an Eastern Conference standings table with a ranked list of where each team in that conference finished in, as well as their head coach, general manager, their wins and losses, whether they made the playoffs, and any notes about discrepancies that appear.

Connected is a table representing an index for all 30 NBA teams that harbors information about their Win/Loss record, the total number of playoff appearances and championships won, the number of head coaches and general managers the team has been through, the current head coach and General Manager (foreign key to the General Managers and Head Coaches tables), and any additional notes.

To go along with this, you have tables of Head Coaches and General Managers, who’s primary keys are foreign keys in the tables displaying NBA standings, the tables representing the team index and the table mentioned above. In the Head Coaches table you have their first and last names, how many teams they coached, and the titles won as a head coach. For the General Managers table, you have the same except that we count the teams they’ve been involved with and championships they won as executives (not necessarily being the General Manager). This is since often these individuals held significant positions in various teams in the season that the team won a championship and should be given credit for it.

Finally, you have two last tables. The first is a table of NBA finals MVPs, with the player’s names, number of MVP awards won, finals MVP awards won and their current team, which is a foreign key to the team index. This is a support to the table of NBA finals results. This lists the season in question, the winning and losing team which are both foreign keys to the Teams index, the ID number of the NBA finals MVP (foreign key), the number of games it took, and the IDs of the coaches and general managers of both the winning and losing teams – foreign keys to the respective tables mentioned above.

It's a good topic for basic SQL since this database can be used for basic trivia questions. If you wanted to get the record of the Milwaukee Bucks in the 2019-20 season as an example, just select the wins and losses columns using a select statement. Since a lot of the tables have foreign keys to other tables, you can use joins to collect more information than is available on a table. If you wanted the name of a General Manager of a team for a specific season, then you could use either a subquery or a join – to be elaborated further later.

I collected the data through hours of research. The information for the NBA standings tables were available through Google. The standings for each year can be shown as the first result that comes up courtesy of the NBA themselves - [click here as an example](https://www.google.ca/search?source=hp&ei=afcxX7nrOcip5wLhpaC4Dg&q=NBA+standings&oq=NBA+standings&gs_lcp=CgZwc3ktYWIQAzIICAAQsQMQgwEyBAgAEAMyCAgAELEDEIMBMggIABCxAxCDATIICAAQsQMQgwEyAggAMgIIADIECAAQAzICCAAyAggAOgsILhCxAxCDARCTAjoICC4QsQMQgwE6BQgAELEDOg0IABCxAxCDARBGEP0BUPgEWMgWYNYXaABwAHgAgAHIA4gBwA2SAQkzLjQuMS4xLjGYAQCgAQGqAQdnd3Mtd2l6&sclient=psy-ab&ved=0ahUKEwj5w_HrgpLrAhXI1FkKHeESCOcQ4dUDCAk&uact=5#sie=lg;/g/11h0yj414v;3;/m/05jvx;st;fp;1;;). The win and loss records, the conference records, and the team ranks were all obtained here. Since the top 8 teams make the playoffs, the top 8 teams would have “Yes” written in the Playoffs column, with the remaining 7 being no. The General Managers and Head Coaches for a specific season were found through Wikipedia, and a website known as RealGM. Wikipedia has pages for each season of an NBA team for the last 10 years at least. This page would harbor generic information about the head coach, general manager, as well as information about the roster, any awards won, the record, in depth player statistics and a summary of the season - [click here as an example](https://en.wikipedia.org/wiki/2018%E2%80%9319_Toronto_Raptors_season). Sometimes the information about the general manager would not be conveniently available in the top right corner of the page. As a result, I went on a website known as RealGM that harbors data for an organization’s executives – [click here for an example](https://basketball.realgm.com/nba/teams/Toronto-Raptors/28/staff-members/Historical/player) – as well as the history of an individual executive – to come in handy later. These were the sources that made up my information about the table standings.

There was also a table for head coaches and general managers. The information for these tables were gathered through Wikipedia, as well as RealGM. These tables served to support the other tables displaying NBA standings and NBA finals results. There was one table known as Teams, which was a table that held information of all 30 teams. The columns of total wins and losses, as well as additional notes about the team’s history were found through Basketball Reference as the team’s “franchise index” being the Google search term. The number of coaches and general managers a team had were found through Wikipedia and RealGM as well. The columns for the NBA finals results were found through Wikipedia, and the supporting table listing NBA finals MVPs were found on Basketball Reference.

There are two discrepancies in this process. The first is that data collected for the NBA standings may not agree with the data found on other websites. Sometimes values were missing for a general manager for a specific season. This was the case for an LA Clippers season somewhere in the midpoint between 2010 and 2020, but for only one season. However, at the time Doc Rivers was president of basketball operations between 2014-2017, so I added his name as a general manager for that season. It is possible that inaccuracy of data could be an issue, as different sites may list different information. The second discrepancy is that this database was made around mid to end July. The 2019-20 season was postponed due to the pandemic, [and is currently resuming in the Orlando bubble](https://en.wikipedia.org/wiki/2019%E2%80%9320_NBA_season). So, the standings and results will most likely be different when this database is being read. Also, since changes often happen in front offices, the values for an NBA team’s current GM or head coach may also change.

SQL Usage

Basic SQL can be used with various tables to get generic information about standing information. If you wanted all 8 playoff teams from the eastern conference for last season you would write one of the following:  
SELECT Team\_Name FROM Eastern\_Conference\_Standings\_2018\_19 WHERE ((Rank >= 1) AND (Rank <= 8));

SELECT Team\_Name FROM Eastern\_Conference\_Standings\_2018\_19 WHERE lower(Playoffs) = “yes”;

If you wanted the wins and losses of your favourite team (in my case the Washington Wizards) from the 2018-19 season then you would type the following:  
SELECT Team\_Name, Wins, Losses FROM Eastern\_Conference\_Standings\_2018\_19 WHERE Team\_Name = “Washington Wizards”;

If you wanted the Id of the General Manager that was at a specific team for a specific season, then you could write the following:  
SELECT GM\_Id FROM Eastern\_Conference\_Standings\_2017\_18 WHERE Team\_Name = “Washington Wizards”;

You could then use this as a subquery to then select the first and last names of the General Manager in the General\_Managers table. Now to the General\_Managers and Head\_Coaches table: you can use select statements to find basic information about a General Manager or head coach. For example, let’s say you wanted to find the number of teams and championships that Gregg Popovich – head coach of the San Antonio Spurs – has won. You would do so like this:  
SELECT Teams\_as\_Head\_Coach, Titles\_Won\_as\_Coach FROM Head\_Coaches WHERE FirstName = “Gregg” AND LastName = “Popovich”;

Doing this for the General Mangers to find the total number of titles Danny Ainge – the general manager of the Boston Celtics – has won being in the front office and the number of team he held as front office, it would look as follows:  
SELECT Teams\_as\_Executive, Titles\_won\_as\_Executive FROM General\_Managers WHERE FirstName = “Danny” AND LastName = “Ainge”;

Now we go to our Teams table. Let’s say that you wanted to find the number of playoff appearances, championships a team has won as well as their conference, say the Washington Wizards. You would type the following:  
SELECT Playoff\_appearances, Championships\_Won, Conference FROM Teams WHERE lower(Team\_Name) = “washington wizards”;

Let’s say you wanted some history about the franchise. You select the additional notes of the franchise like so:  
SELECT Additional\_Notes FROM Teams WHERE lower(Team) = “washington wizards”;

Lastly, let’s say you had a theory that organizations with more head coaches and general managers – thus more turnover – would make less playoff appearances and win less titles. You could do this by selecting the rows of Playoff\_appearances, Championships\_Won from teams but ordering by the sum of each team’s playoff appearances and championships won. One thing to note is that teams were established at different times, so there will be no correlation here that helps since the longer a team is in business, the more head coaches and general managers it will have.

SELECT Playoff\_appearances, Championships\_Won FROM Teams ORDER BY (Total\_Coaches + Total\_GMs);

You could order the number of teams by the total number of games they played as well:  
SELECT Team (Total\_Wins + Total\_Losses) AS Total\_Games\_Played FROM Teams ORDER BY (Total\_Wins + Total\_Losses) DESC;

You could order the teams by their win percentage:  
SELECT ((Total\_Wins \* 1.0) / ((Total\_Wins \* 1.0) + (Total\_Losses \* 1.0))) AS Win\_Pct, Team FROM Teams ORDER BY Win\_Pct DESC;

If you wanted the id numbers of either the current head coach or the current general manager of a specific team, you could simply select them from the database:  
SELECT Current\_GM\_ID, Current\_Coach\_ID FROM Teams WHERE lower(Team) = “washington wizards”;

Now we go to the Finals MVP table: if we wanted to select the name of a player with a certain id number, then could do the following as an example:  
SELECT FirstName, LastName FROM Finals\_MVPs WHERE Player\_id = 1;

If we wanted to find the accolades of a specific player, then we would do the following as an example:  
SELECT All\_Star\_Appearances, MVP\_Awards, Finals\_MVP\_Awards FROM Finals\_MVPs WHERE lower(FirstName) = “kawhi” AND lower(LastName) = “leonard”;

Now we go the NBA\_Finals\_Results table. Let’s say we wanted to find out which team won and which one lost for a specific season. For the 2010-11 NBA season, we could put in the following:  
SELECT Winner, Loser FROM NBA\_Finals\_Results WHERE Season = “2010-11”;

You could replace the columns with Finals\_MVP\_ID to get the ID of the Finals\_MVP, the Number\_of\_games to get the number of games it took, the Winner\_Coach\_ID to get the coach’s ID for the winning team, and the Winner\_GM\_ID to get the General Manager’s ID for the winning team.

Now we go to advanced SQL. We start with the NBA\_Finals\_Results table. Here we can use subqueries from the Finals MVP table to select certain rows from the table storing finals results. If we wanted to find the finals series where Kawhi Leonard was the Finals MVP, we can do so as follows:  
SELECT \* FROM NBA\_Finals\_Results WHERE Finals\_MVP\_ID IN (SELECT Player\_Id FROM Finals\_MVPs WHERE FirstName = “Kawhi” AND LastName = “Leonard”);

If we wanted to set constraints as to giving us back rows from the Finals Results table where the Finals MVP fit a certain criteria of accolades in the Finals\_MVPs table, we could do the following:  
SELECT \* FROM NBA\_Finals\_Results WHERE Finals\_MVP\_ID IN (SELECT Player\_Ids FROM Finals\_MVPs WHERE Finals\_MVPs > 1 AND MVP\_Awards >= 1);

Since we also have the Teams index connected to the Finals MVP table, we can select based off the Teams index and the Finals MVP table. If we wanted results from the finals where the Finals MVP had at more than one Finals MVP award and the coach is still with the same team we would do the following:  
SELECT \* FROM NBA\_Finals\_Results WHERE Finals\_MVP\_ID IN (SELECT Player\_id FROM Finals\_MVPs WHERE Finals\_MVP\_Awards > 1) AND Winner\_Coach\_ID IN (SELECT Current\_Coach\_ID FROM Teams);

If we wanted to grab columns from both tables, then we can join both tables on the Finals\_MVP\_ID of the NBA\_Finals\_Results table and the Player\_Id column of the Finals\_MVPs table. We could select columns from both tables, such as the winning, team, the finals mvp, and how many MVPs they won:  
SELECT NBA\_Finals\_Results.Winner, Finals\_MVPs.First\_Name, Finals\_MVPs.Last\_Name, Finals\_MVPs.MVP\_Awards FROM NBA\_Finals\_Results JOIN Finals\_MVPs ON Finals\_MVPs.Player\_Id = NBA\_Finals\_Results.Finals\_MVP\_ID;

We use these subqueries for the connecting tables to select rows based on any number of predetermined criteria, such as which titles were won where the Finals MVP is still on the same team. But now we go over to the NBA standings tables. So from here you could use joins with the standings for the same conference to get all 16 teams that made the playoffs. So for the 2019-20 NBA season we could do:

SELECT Eastern\_Conference\_Standings\_2019\_20.Team\_Name, Western\_Conference\_Standings\_2019\_20.Team\_Name FROM Eastern\_Conference\_Standings\_2019\_20 JOIN Western\_Conference\_Standings\_2019\_20 ON Western\_Conference\_Standings\_2019\_20.Team\_Rank = Eastern\_Conference\_Standings\_2019\_20.Team\_Rank WHERE lower(Eastern\_Conference\_Standings\_2019\_20.Playoffs) = “yes” AND lower(Western\_Conference\_Standings\_2019\_20.Playoffs) = “yes”;

This will put it in separate lists. We could put it in one list like so using subqueries:

SELECT Teams.Team FROM Teams WHERE Teams.Team IN  
(SELECT Eastern\_Conference\_Standings\_2019\_20.Team\_Name FROM Eastern\_Conference\_Standings\_2019\_20 WHERE Eastern\_Conference\_Standings\_2019\_20.Playoffs = "Yes") OR Teams.Team IN  
(SELECT Western\_Conference\_Standings\_2019\_20.Team\_Name FROM Western\_Conference\_Standings\_2019\_20 WHERE Western\_Conference\_Standings\_2019\_20.Playoffs = "Yes");

Once we join the two tables on their Team\_Rank attributes, we select other attributes based on other criteria, such as the id values of the general managers with playoff teams, or the id values of head coaches who made the playoffs. You can use these in subqueries to determine which coaches coached playoff teams, or which General Managers supposedly made the right trades to put their teams in the playoffs. The respective codes would look like so:

SELECT General\_Managers.FirstName AS GM\_Fname, General\_Managers.LastName AS GM\_Lname FROM General\_Managers WHERE General\_Managers.GM\_Id IN   
(SELECT GM\_Id FROM Western\_Conference\_Standings\_2019\_20 WHERE Playoffs = "Yes") OR General\_Managers.GM\_Id IN   
(SELECT GM\_Id FROM Eastern\_Conference\_Standings\_2019\_20 WHERE Playoffs = "Yes");

SELECT Head\_Coaches.FirstName AS Coach\_Fname, Head\_Coaches.LastName AS Coach\_Lname FROM Head\_Coaches WHERE Head\_Coaches.Coach\_Id IN   
(SELECT Coach\_Id FROM Western\_Conference\_Standings\_2019\_20 WHERE Playoffs = "Yes") OR Head\_Coaches.Coach\_Id IN  
(SELECT Coach\_Id FROM Eastern\_Conference\_Standings\_2019\_20 WHERE Playoffs = "Yes");

You can also join the NBA Teams index table on the eastern conference standings table to get more information about a team in the eastern conference:

SELECT Eastern\_Conference\_Standings\_2019\_20.Team\_Rank, Eastern\_Conference\_Standings\_2019\_20.Team\_Name, Teams.Playoff\_appearances, Teams.Championships\_Won, Teams.Additional\_Notes FROM Teams JOIN Eastern\_Conference\_Standings\_2019\_20 ON Teams.Team = Eastern\_Conference\_Standings\_2019\_20.Team\_Name;

We can also use the Teams table and the General Managers and Head Coaches tables to determine the staff who currently have a job in the NBA. The according code to find the names of the General Managers and coaches are:

SELECT FirstName, LastName FROM General\_Managers WHERE GM\_Id IN (SELECT Current\_GM\_ID FROM Teams);

SELECT FirstName, LastName FROM Head\_Coaches WHERE Coach\_Id IN (SELECT Current\_Coach\_ID FROM Teams);

Normal Form Verification

Now we verify the normality of all our relations. Let’s start with the Finals\_MVPs table:

R = (Player\_id, FirstName, LastName, All\_Star\_Appearances, MVP\_Awards, Finals\_MVP\_Awards, Current\_Team)  
K = Player\_id  
FD1 = FirstName, LastName -> MVP\_Awards  
FD2 = FirstName, LastName -> Finals\_MVP\_Awards  
FD3 = FirstName, LastName -> Current\_Team

Since the primary key is only one attribute, by the trick definition we do not need to apply the 2NF test since it’s already in 2NF. There are no transitive dependencies, so as a result this relation is in 3NF. I will assume that for the most part, the players’ names will be unique to them, and so the awards and current will be dependent on the name – not certain in theory.

Now the NBA\_Finals\_Results table:

R = (Season, Winner, Loser, Finals\_MVP\_ID, Number\_of\_games, Winner\_Coach\_ID, Winner\_GM\_ID, Losing\_Coach\_ID, Losing\_GM\_ID)  
K = Season   
FD1 = Winner -> Finals\_MVP\_ID  
FD2 = Winner -> Winner\_Coach\_ID  
FD3 = Winner -> Winner\_GM\_ID  
FD4 = Loser -> Losing\_Coach\_ID  
FD5 = Loser -> Losing\_GM\_ID

Since the primary key is one attribute, it’s in 2NF by the trick definition. There are no transitive dependencies, so this relation is in 3NF. We can say that there are no functional dependencies between the winning and losing teams and the specific season because the season doesn’t necessarily dictate who will be in the finals.

Now to the Head\_Coaches table:

R = (Coach\_Id, FirstName, LastName, Teams\_as\_Head\_Coach, Titles\_Won\_as\_Coach)  
K = Coach\_Id  
FD1 = (FirstName, LastName) -> Teams\_as\_Head\_Coach  
FD2 = (FirstName, LastName) -> Titles\_Won\_as\_Coach

There is only one primary key attribute, so it’s in 2NF by the trick definition. There are no transitive dependencies, so this relation is in 3NF. The functional dependencies exist above because I assume that most of the time the coach’s name will be unique to them, and thus the teams they’ve been on and the titles won by them will be unique to that coach’s name – but this not certain in theory.

Now to the General\_Managers table:

R = (GM\_Id, FirstName, LastName, Teams\_as\_Executive, Titles\_won\_as\_Executive)  
K = GM\_Id  
FD1 = (FirstName, LastName) -> Teams\_as\_Exxecutive  
FD2 = (FirstName, LastName) -> Titles\_won\_as\_Executive

There is only one attribute that is the primary key. So by the trick definition it is in 2NF. Since there are no transitive dependencies, this relation is in 3NF. These functional dependencies are there because the executive’s name is generally what determines the number of teams and titles won by that executive, since there is a good chance – but not certain in theory – that the name is unique to them.

Now to the Teams table:

R = (Team, Total\_Wins, Total\_Losses, Playoff\_appearances, Championships\_Won, Total\_Coaches, Total\_GMs, Current\_Coach\_ID, Current\_GM\_ID, Conference, Additional\_Notes)  
K = Team  
FD1 = Team -> Total\_Wins  
FD2 = Team -> Total\_Losses  
FD3 = (Total\_Wins, Total\_Losses) -> Playoff\_appearances  
FD4 = Playoff\_appearances -> Championships\_Won  
FD5 = Team -> Total\_Coaches  
FD6 = Team -> Total\_GMs  
FD7 = Team -> Conference  
FD8 = Team -> Additional\_Notes

Different teams were established in different times, thus affecting the total games played and hence partially affecting the Win and Loss totals. This can also affect how many General Managers and Head Coaches have gone in and out of the organization. There have been situations where a coach or a general manager stays as part of an organization’s legacy; Gregg Popovich in San Antonio for example. However, this is rare so I will refrain from having functional dependencies between teams and their current coach or general manager. The relation is in 2NF by the trick definition because of only one primary key. However, because there are transitive dependencies between FD1, FD2, FD3 and FD4 this table is not in 3NF. There is also no use trying to bring it into 3NF because since teams were established at different times, they will have made more playoff appearances and won more championships because there were fewer teams in the league. The Celtics are an example. Also, removing the number of playoff appearances may take away from the purpose of the database.

Now we go to the standings tables. All of them are built the same, just with different titles, so one example will do:

R = (Team\_Rank, Team, Wins, Losses, Coach\_Id, GM\_Id, Playoffs, Notes)  
K = Team\_Rank  
FD1 = Team\_Rank -> Playoffs  
FD2 = (Wins, Losses) -> Notes  
FD3 = Team -> Coach\_Id  
FD4 = Team -> GM\_Id  
FD5 = Team\_Rank -> Notes  
FD6 = Team -> Notes

I have that FD3, FD5, and FD6 all have Notes on the right end because the notes for the team’s regular season can be dependant on several things. Team wins and losses can affect the content of the notes, because often teams have the same record but there are other factors that go into the team rank. There were multiple cases of teams having the same win/loss record but one being ranked higher. As a result, there isn’t a functional dependency between Team\_Rank and (Wins, Losses). Sometimes things happen with a team’s staffing changes, where if you look online a coach was there for only a part of the season. If they’re there for longer their ID is in the table for that season, hence why Notes partially depend on the Team column. The relation is in 2NF because of there being one attribute for the primary key and the trick definition. Since there are no transitive dependencies – or not enough of a dependency between Team\_Rank and (Wins, Losses) to be transitive – this relation is in 3NF. Thus all tables listing NBA standings are in 3NF.

References:

1. Main Page. (2020, July 23). Retrieved August 13, 2020, from <https://www.wikipedia.org/>
2. Basketball News, Rumors, Scores, Stats, Analysis, Depth Charts, Forums. (n.d.). Retrieved August 13, 2020, from <https://basketball.realgm.com/>
3. Basketball Statistics and History. (n.d.). Retrieved August 13, 2020, from <https://www.basketball-reference.com/>