

NBA DATABASE

CSS 475 Winter 2019 – Team Avatar

Jeremy Tandjung, Nathan Phan, Aaron Handjojo, Brandon Posey

Contents

1. INTRODUCTION	2
2. Application.....	2
3. Entities	2
4. Design	4
4.1. Assumptions	4
4.2. Entity Relationship (ER) Diagram	5
Figure 1: NBA ER Diagram	6
4.3. Relational Model	7
Figure 2: NBA RM Diagram	8
4.4. Constraints.....	8
5. Queries	13
6. Data Generation	14
7. SQL Script	14
7.1. Methodology	14
7.2. Database Creation	14
7.3. Table Creation	15
7.4. Table Population	16
7.5. Sample Queries	18
8. Schedule.....	19
8.1. Team Formation (01-09-2019).....	19
8.2. Diagram (01-21-2019).....	19

8.3.	Tooling Section (01-21-2019).....	19
8.4.	Develop Database (03-08-2019)	20
8.5.	Finish Database and Presentation (03-18-2019)	20
9.	Work Distribution	20
10.	Tooling Assessment	20
10.1.	DBMS: MySQL Workbench CE.....	21
10.2.	UI: HTML/CSS	21
10.3.	Version Control: GitHub	21
10.4.	Web HOSTING Service: AWS or USBWebServer	21

1. INTRODUCTION

Team Avatar intends to create a basketball database based on the properties of the National Basketball Association. The database will track team and player information, statistics, awards, etc within the current NBA season, which is the 2018-19 NBA season.

Team Avatar will implement an update meeting cycle, where every week we meet up to give each other an update on what we have done in the past week.

2.APPLICATION

The application is primarily used by NBA teams to determine whether a player is playing well or not. The database could also be used by fantasy league users to see if their current fantasy roster is playing well or not. NBA fans could also look up trivial information of their favorite team and players, such as birthdate, age, height, and many more.

3.ENTITIES

The following are the entity data types that are in the database

- **TEAMS:** The 30 Teams that are in the NBA as of the current NBA season; 2018-19 NBA Season
 - TeamID
 - Name
 - CoachID
 - ArenaID
 - Conference
 - Division
 - Championships
 - Wins
 - Losses
- **ARENAS:** The current and past arenas used by the 30 NBA teams.
 - ArenaID
 - Name
 - Capacity
 - Address
 - City
 - State
- **HEAD_COACHES:** The current and past head coaches of the 30 NBA teams
 - CoachID
 - Name
 - Wins
 - Losses
 - Dob
- **PLAYER:** The current roster of the 30 NBA teams
 - PlayerID
 - Name
 - TeamID
 - FName (First Name)
 - LName (Last Name)
 - DOB (Date of Birth)

- Height
 - Weight
 - DraftYear (The Year the player got drafted)
- STATISTICS: The player's per game statistics
 - PlayerID
 - StatsID
 - PPG (Points Per Game)
 - APG (Assist Per Game)
 - RPG (Rebound Per Game)
 - SPG (Steals Per Game)
 - BPG (Blocks Per Game)
 - FG% (Field Goal Percentage)
 - 3FG% (3-Point Field Goal Percentage)
- AWARDS: The awards given to the players for their performance in the regular season
 - AwardID
 - Name
 - Description

4. DESIGN

In this section, we will discuss the assumption, design decision, and both ER and RM diagrams that our group have chosen to work it

4.1. ASSUMPTIONS

1. This database only contains data from the current NBA season during the making of this project, which is the 2018-19 NBA season.
2. All player statistics and information are up to date right before the NBA All-Star Weekend which is Feb 16th, 2019.
3. This database will include all current and past arenas that are used by the 30 NBA teams.

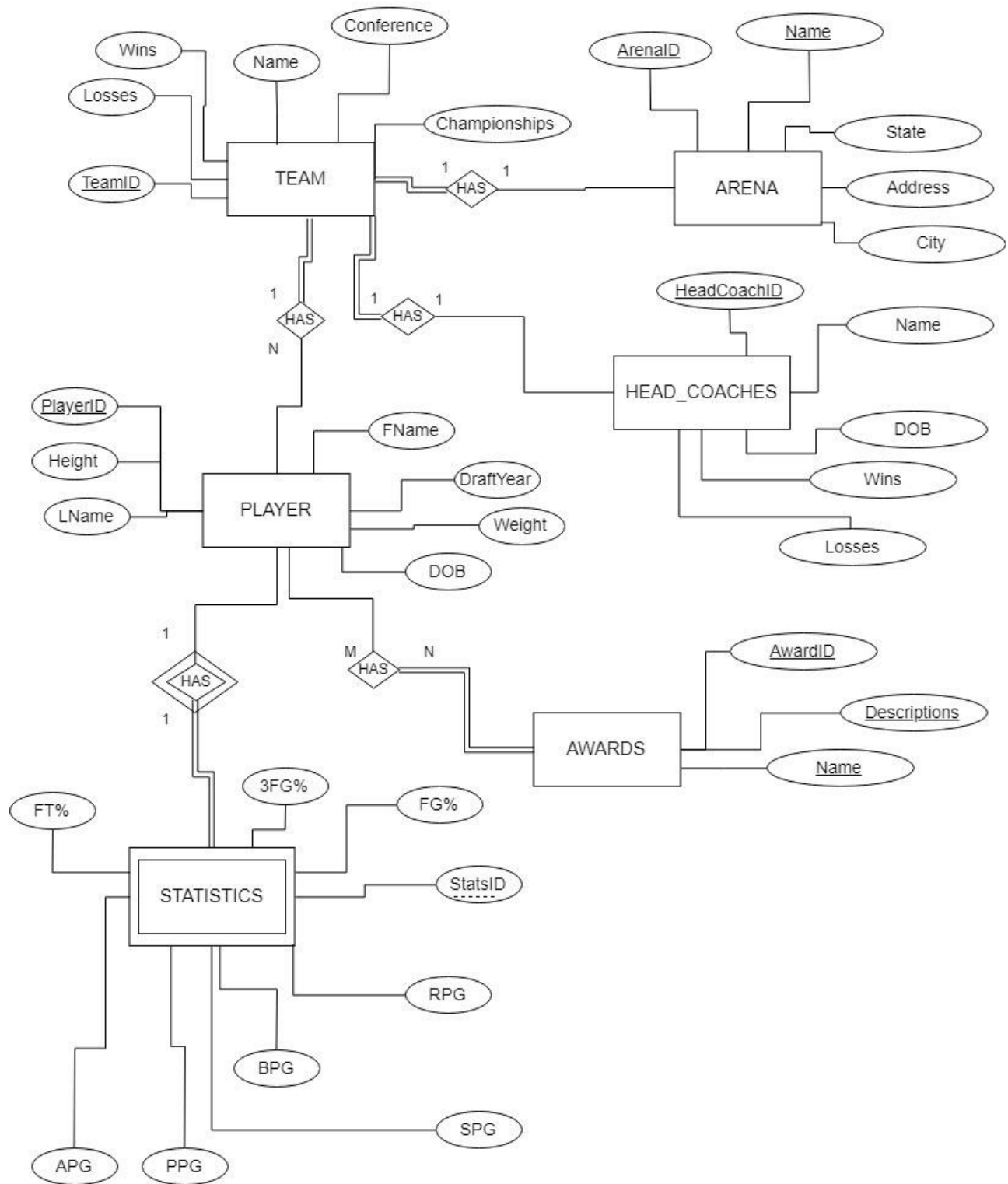
4. This database will include past and current head coaches that has coached an NBA team. That means, coaches who are currently not coaching an NBA team are included in the database.

4.2. ENTITIY RELATIONSHIP (ER) DIAGRAM

The Entity Relationship Diagram provides an abstract overview of the various relationships between the entities derived from the requirements of the NBA Database. The strong entities include the following entities: TEAM, ARENA, PLAYER, HEADCOACHES, and AWARDS. The only weak entity is STATISTICS, which has an identifying relationship with the entity PLAYER. The attributes that correspond with each entity are clearly connected. In addition, attributes that serve as keys are underlined to indicate their status as candidate keys, or partially underlined to indicate their status as partial keys. Each relationship between entities has its cardinality and participation status clearly labeled. Refer to Figure 1 for further details regarding the diagram.

Team Avatar chose to create “HAS” relationships between TEAM and the entities of ARENA, HEAD_COACHES, and PLAYER. The reasoning for this decision is that TEAMS in the NBA have direct relationships with those entities wherein the latter entities are identified as belonging to the TEAM. Since an NBA team needs to have a relationship with those entities, it was decided to impose total participation on the TEAM side of those relationships. In addition, the cardinality ratio of the relationships are 1:1 in the case of TEAM and ARENA, and TEAM, and HEAD_COACHES

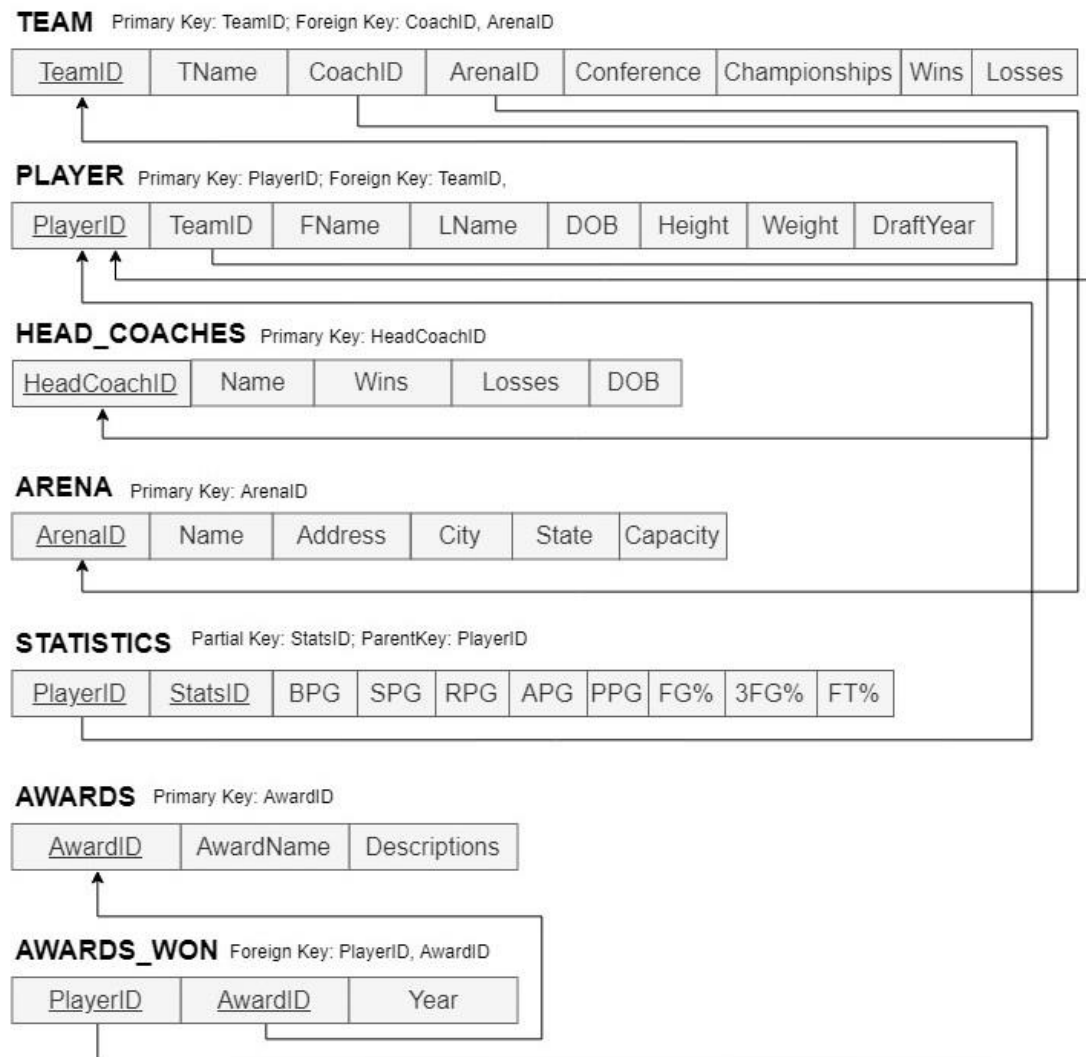
Figure 1: NBA ER Diagram



4.3. RELATIONAL MODEL

The Relational Model provides an abstract overview of the different relations derived from the requirements. Each relation will include its entity's attributes including keys such as primary keys, foreign keys, and partial keys. The keys that are referenced from other relations are modeled using arrows to demonstrate the origin of the foreign key. In this case, the relation AWARDS_WON represents the relationship between PLAYER and AWARDS. The relation itself will have no distinctive key as the relationship between PLAYER and AWARDS with the foreign keys of PlayerID, AwardID, and the relation's attribute of Year providing meaning and uniqueness for the AWARDS_WON relation. Refer to Figure 2 for further details regarding the Relational Model schema.

Figure 2: NBA RM Diagram



4.4. CONSTRAINTS

- TEAM
 - TeamID:
 - PRIMARY KEY
 - Not NULL

- Domain: All positive integers
- TName:
 - Not NULL
 - Domain: VarChar(40)
- Conference:
 - Not NULL
 - Domain: VarChar(40)
- Championship:
 - Not NULL
 - Domain: All non-negative integers
- Wins:
 - Not NULL
 - Domain: All non-negative integers
- Loses:
 - Not NULL
 - Domain: All non-negative integers
- CoachID:
 - FOREIGN KEY from HEAD_COACHES
 - Not NULL
 - Domain: All non-negative integers
- ArenaID:
 - FOREIGN KEY from ARENA
 - Not NULL
 - Domain: All non-negative integers
- PLAYER
 - PlayerID:
 - PRIMARY KEY
 - Not NULL
 - Domain: All positive integers
 - FName:
 - Not NULL

- Domain: All non-numerical VarChar(40)
- LName:
 - Not NULL
 - Domain: All non-numerical VarChar(40)
- DOB:
 - Not NULL
 - Domain: integers in the format mm/dd/yy where: $1 \leq mm \leq 12$, $1 \leq dd \leq 31$ (for months with 31 days) or $1 \leq dd \leq 30$ (for months with 30 days) or $1 \leq dd \leq 28$ (for February in a non-leap year) or $1 \leq dd \leq 29$ (for February in a leap year), and yy being the last two digits of years between 1900 and 2099.
- Height:
 - Not NULL
 - Domain: All non-negative integers (in cm)
- Weight:
 - Not NULL
 - Domain: All non-negative integers (in kg)
- DraftYear:
 - Not NULL
 - Domain: 4-digit positive integer between 1900 and 2099
- TeamID:
 - FOREIGN KEY from TEAM
 - Not NULL
 - Domain: All non-negative integers
- HEAD_COACHES
 - HeadCoachID:
 - PRIMARY KEY
 - Not NULL
 - Domain: All positive integers
 - Name:
 - Not NULL
 - Domain: All non-numerical VarChar(40)
 - Wins:

- Not NULL
 - Domain: All non-negative integers
- Loses:
 - Not NULL
 - Domain: All non-negative integers
- DOB:
 - Not NULL
 - Domain: integers in the format mm/dd/yy where: $1 \leq \text{mm} \leq 12$, $1 \leq \text{dd} \leq 31$ (for months with 31 days) or $1 \leq \text{dd} \leq 30$ (for months with 30 days) or $1 \leq \text{dd} \leq 28$ (for February in a non-leap year) or $1 \leq \text{dd} \leq 29$ (for February in a leap year), and yy being the last two digits of years between 1900 and 2099.
- ARENA
 - ArenaID:
 - PRIMARY KEY
 - Not NULL
 - Domain: All positive integers
 - Name:
 - Not NULL
 - Domain: All non-numerical VarChar(40)
 - Address:
 - Not NULL
 - Domain: VarChar(40)
 - City:
 - Not NULL
 - Domain: VarChar(40) of a name of a valid city in the US within the State.
 - State:
 - Not NULL
 - Domain: Char(2) of a name of a US State in abbreviation
 - Capacity:
 - Not NULL

- Domain: All non-negative integers
- STATISTICS
 - StatsID:
 - PARTIAL KEY
 - Not NULL
 - Domain: All positive integers
 - PlayerID:
 - PARENT KEY from PLAYER
 - Not NULL
 - Domain: All positive integers
 - BPG:
 - Not NULL
 - Domain: All non-negative double
 - SPG:
 - Not NULL
 - Domain: All non-negative double
 - RPG:
 - Not NULL
 - Domain: All non-negative double
 - APG:
 - Not NULL
 - Domain: All non-negative double
 - PPG:
 - Not NULL
 - Domain: All non-negative double
 - FG%:
 - Not NULL
 - Domain: All non-negative double
 - 3FG%:
 - Not NULL
 - Domain: All non-negative double

- FT%:
 - Not NULL
 - Domain: All non-negative double
- AWARDS
 - AwardID:
 - PRIMARY KEY
 - Not NULL
 - Domain: All positive integers
 - AwardName:
 - Not NULL
 - Domain: VarChar(40)
 - Description:
 - Not NULL
 - Domain: VarChar(400)
- AWARDS_WON
 - AwardID:
 - PARENT KEY from AWARDS
 - Not NULL
 - Domain: All positive integers
 - PlayerID:
 - PARENT KEY from PLAYER
 - Not NULL
 - Domain: All positive integers
 - Year:
 - Not NULL
 - Domain: 4-digit positive integer between 1900 and 2099

5. QUERIES

The NBA database will allow any interested party, namely NBA fantasy players, to query information about players and teams to see trends and make predictions about team, and individual performance for the 2018-2019 season.

For example, an NBA fantasy player may query the database to search for every player who currently plays for the Boston Celtics. The Fantasy player may then search for players who are not listed as out or suspended for the season.

The Fantasy player can then search the database with the existing filters to determine which of the players have a higher fantasy point average. The fantasy player can also narrow down the results by querying the database to search for players who have a higher than average projected points for the year.

6. DATA GENERATION

Team Avatar will use public information provided by basketball-reference.com to populate the database with NBA teams and players. The statistics that are available for each team and player will populate the team and player records. Team Avatar will use basketball-reference.com as their primary source of data.

7. SQL SCRIPT

As mentioned before, team Avatar uses MySQL as our DBMS to support our NBA database. We have attached a .sql file for all the script below in another separate file.

7.1. METHODOLOGY

Team Avatar used <http://basketball-reference.com> as our main source for data inputs. They have convenient features that lets us download players', teams', and coaches' information in an excel format. After that we arranged the data so that they fit to our tables' attributes, then we converted the excel files into csv files. One of our members then wrote a python program that converts the csv file onto a SQL script for inserting values onto our tables.

7.2. DATABASE CREATION

```
DROP DATABASE IF EXISTS css475_NBA;  
CREATE DATABASE css475_NBA;  
USE css475_NBA;
```

7.3. TABLE CREATION

```
• DROP TABLE IF EXISTS TEAM;
• CREATE TABLE TEAM
(
    TeamId          INT unsigned PRIMARY KEY auto_increment,
    Name            VARCHAR(40),
    CoachId         INT unsigned,
    ArenaId         INT unsigned,
    Conference       ENUM('West', 'East'),
    Division        ENUM('Pacific', 'Northwest', 'Southwest', 'Atlantic', 'Central', 'Southeast'),
    Championships   INT
);

• DROP TABLE IF EXISTS PLAYER;
• CREATE TABLE PLAYER
(
    PlayerId        INT unsigned NOT NULL PRIMARY KEY auto_increment,
    FName           VARCHAR(40),
    LName           VARCHAR(40),
    TeamId          INT unsigned,
    DOB             DATE,
    Height_ft       INT unsigned,
    Height_inch     INT unsigned,
    Weight          INT unsigned,
    DraftYear       YEAR
);
```

```
• DROP TABLE IF EXISTS ARENA;
• CREATE TABLE ARENA
(
    ArenaId         INT unsigned NOT NULL PRIMARY KEY auto_increment,
    Name            VARCHAR(50),
    Address         VARCHAR(80),
    City            VARCHAR(40),
    State           VARCHAR(2),
    Capacity        INT unsigned,
    BuiltIn         INT
);

• DROP TABLE IF EXISTS HEAD_COACH;
• CREATE TABLE HEAD_COACH
(
    CoachId         INT unsigned NOT NULL PRIMARY KEY,
    Name            VARCHAR(40),
    Wins            INT,
    Losses          INT,
    DOB             DATE
);
```

- DROP TABLE IF EXISTS STATS;
- CREATE TABLE STATS
 - (

PlayerId	INT unsigned,
StatsId	INT unsigned,
PPG	DECIMAL(5,2),
APG	DECIMAL(5,2),
RPG	DECIMAL(5,2),
BPG	DECIMAL(5,2),
SPG	DECIMAL(5,2),
FG	DECIMAL(5,2),
FG3	DECIMAL(5,2),

 CONSTRAINT
 FOREIGN KEY (PlayerId)
 REFERENCES PLAYER(PlayerId)
 -);
- DROP TABLE IF EXISTS AWARD;
- CREATE TABLE AWARD
 - (

AwardId	INT unsigned,
Name	VARCHAR(40),
Description	VARCHAR(100)

)
 -);

7.4. TABLE POPULATION

Below are the SQL script and the data inputs for TEAM, HEAD_COACH, and ARENA.

The data shown here are not the complete list as the data was too much to fit in this document.

```
INSERT INTO TEAM
VALUES
  (1, 'Philadelphia 76ers', 1026, 29, 'East', 3, 34, 20),
  (2, 'Portland Trail Blazers', 1028, 14, 'West', 1, 33, 21),
  (3, 'Milwaukee Bucks', 1019, 10, 'East', 1, 40, 13),
  (4, 'Chicago Bulls', 1005, 27, 'East', 6, 12, 42),
  (5, 'Cleveland Cavaliers', 1008, 17, 'East', 1, 11, 43),
  (6, 'Boston Celtics', 1002, 25, 'East', 17, 35, 20),
  (7, 'Los Angeles Clippers', 1015, 21, 'West', NULL, 30, 26),
  (8, 'Memphis Grizzlies', 1017, 9, 'West', NULL, 22, 34),
  (9, 'Atlanta Hawks', 1001, 22, 'East', 1, 18, 36),
  (10, 'Miami Heat', 1018, 1, 'East', 3, 25, 27),
```



```
INSERT INTO HEAD_COACH
VALUES
```

```
(1001, 'Lloyd Pierce', 22, 43, '1976-05-11'),
(1002, 'Brad Stevens', 259, 215, '1976-10-22'),
(1003, 'Kenny Atkinson', 81, 149, '1967-06-02'),
(1004, 'Fred Hoiberg', 115, 155, '1972-10-15'),
(1005, 'Jim Boylen', 13, 27, '1965-04-18'),
(1006, 'James Borrego', 39, 54, '1977-11-12'),
(1007, 'Tyronn Lue', 128, 83, '1977-05-03'),
(1008, 'Larry Drew', 159, 211, '1958-04-02'),
(1009, 'Rick Carlisle', 745, 614, '1959-10-27'),
(1010, 'Mike Malone', 200, 215, '1971-01-01'),
(1011, 'Dwane Casey', 404, 338, '1957-04-17'),
(1012, 'Steve Kerr', 309, 82, '1965-09-27'),
(1013, 'Mike D\'Antoni', 613, 495, '1951-05-08'),
(1014, 'Nate McMillan', 609, 549, '1964-08-03'),
(1015, 'Doc Rivers', 883, 653, '1961-10-13'),
(1016, 'Luke Walton', 91, 137, '1980-03-28'),
(1017, 'J.B. Bickerstaff', 77, 122, '1979-03-10'),
(1018, 'Erik Spoelstra', 513, 354, '1970-11-01'),
(1019, 'Mike Budenholzer', 261, 213, '1969-08-06'),
(1020, 'Tom Thibodeau', 352, 246, '1958-01-17'),
(1021, 'Ryan Saunders', 10, 13, '1986-04-28'),
(1022, 'Alvin Gentry', 477, 540, '1954-11-05'),
```

```
INSERT INTO ARENA
VALUES
```

```
(1, 'American Airlines Arena', '601 Biscayne Boulevard', 'Miami', 'FL', 19600, 1999),
(2, 'American Airlines Center', '2500 Victory Avenue', 'Dallas', 'TX', 19200, 2001),
(3, 'Amway Center', '400 West Church Street', 'Orlando', 'FL', 18846, 2010),
(4, 'AT&T Center', '1 AT&T Center Parkway', 'San Antonio', 'TX', 18418, 2002),
(5, 'Bankers Life Fieldhouse', '125 South Pennsylvania Street', 'Indianapolis', 'IN', 17923, 1999),
(6, 'Barclays Center', '620 Atlantic Avenue', 'Brooklyn', 'NY', 17732, 2012),
(7, 'Capital One Arena', '601 F Street NW', 'Washington D.C.', 'DC', 20356, 1997),
(8, 'Chesapeake Energy Arena', '100 West Reno Avenue', 'Oklahoma City', 'OK', 18203, 2002),
(9, 'FedExForum', '191 Beale Street', 'Memphis', 'TN', 17794, 2004),
(10, 'Fiserv Forum', '1111 Vel R. Phillips Avenue', 'Milwaukee', 'WI', 17500, 2018),
(11, 'Golden 1 Center', '500 David J. Stern Walk', 'Sacramento', 'CA', 17583, 2016),
(12, 'Little Caesars Arena', '2645 Woodward Avenue', 'Detroit', 'MI', 20491, 2017),
(13, 'Madison Square Garden', '4 Pennsylvania Plaza', 'New York City', 'NY', 19812, 1968),
(14, 'Moda Center', '1 Center Court', 'Portland', 'OR', 19441, 1995),
(15, 'Oracle Arena', '7000 Coliseum Way', 'Oakland', 'CA', 19596, 1966),
(16, 'Pepsi Center', '1000 Chopper Circle', 'Denver', 'CO', 19520, 1999),
(17, 'Quicken Loans Arena', '1 Center Court', 'Cleveland', 'OH', 20562, 1994),
(18, 'Scotiabank Arena', '40 Bay Street', 'Toronto', 'ON', 19800, 1999),
(19, 'Smoothie King Center', '1501 Dave Dixon Drive', 'New Orleans', 'LA', 16687, 1999),
(20, 'Spectrum Center', '333 East Trade Street', 'Charlotte', 'NC', 19077, 2005),
(21, 'Staples Center', '1111 South Figueroa Street', 'Los Angeles', 'CA', 19068, 1999),
(22, 'State Farm Arena', '1 State Farm Drive', 'Atlanta', 'GA', 18118, 1999),
(23, 'Talking Stick Resort Arena', '201 E Jefferson St', 'Phoenix', 'AZ', 18055, 1992),
(24, 'Target Center', '600 N First Avenue North', 'Minneapolis', 'MN', 18978, 1990),
(25, 'TD Garden', '100 Legends Way', 'Boston', 'MA', 18624, 1995),
```

7.5. SAMPLE QUERIES

```
219 #Teams that have 5 championships or more
220 • SELECT TEAM.Name,
221      Championships
222 FROM TEAM
223 WHERE Championships >= 5;
```

Result Grid | Filter Rows: | Export: | Wrap

Name	Championships
Chicago Bulls	6
Boston Celtics	17
Los Angeles Lakers	16
San Antonio Spurs	5
Golden State Warriors	6

```
225 #Search a team that has Luke Walton as their head coach
226 • SELECT TEAM.Name
227 FROM TEAM
228      JOIN HEAD_COACH ON HEAD_COACH.CoachId = TEAM.CoachId
229 WHERE HEAD_COACH.Name = 'Luke Walton';
```

```
231 #Display all players sorted by height
232 • SELECT CONCAT(Fname, ' ', Lname) AS Player,
233      CONCAT(Height_ft, '\'', Height_inch, '\'') AS Height
234 FROM PLAYER
235 ORDER BY Height_ft DESC, Height_inch DESC;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

Player	Height
Boban Marianovic	7'3"
Mevers Leonard	7'1"
Joel Embiid	7'0"
Justin Patton	7'0"
Zach Collins	7'0"
Jusuf Nurkic	7'0"
Enes Kanter	6'11"
Skal Labissiere	6'11"
Jonah Bolden	6'10"
Ben Simmons	6'10"
Tobias Harris	6'9"
Amir Johnson	6'9"

7.6. CONSTRAINT VIOLATION TEST CASE

```
CREATE TABLE ARENA
(
    ArenaId      INT unsigned NOT NULL PRIMARY KEY auto_increment,
    Name         VARCHAR(50) NOT NULL,
    Address      VARCHAR(80) NOT NULL,
    City         VARCHAR(40) NOT NULL,
    State        VARCHAR(2) NOT NULL,
    Capacity     INT unsigned NOT NULL,
    BuiltIn      INT NOT NULL
);
```

```
INSERT INTO ARENA
VALUES
(1, 'American Airlines Arena', '601 Biscayne Boulevard', 'Miami', 'FL', 19600, NULL),
```

20 22:45:15 INSERT INTO ARENA VALUES (1, 'American Airlines Arena', '601 Biscayne Boulevard', 'Miami', 'FL', 19600,... Error Code: 1048. Column 'BuiltIn' cannot be null

8. SCHEDULE

Milestones for this project largely mirror the project iteration for each of the following assignments. Dates shown are the date it was finished on or the expected deadline.

8.1. TEAM FORMATION (01-09-2019)

1. Set team by laws
2. Brainstorm project ideas
3. Check feasibility
4. Decide project idea

8.2. DIAGRAM (01-21-2019)

1. Create Entity Relationship Diagram
2. Create Relational Model Schema

8.3. TOOLING SECTION (01-21-2019)

1. Choose DBMS
2. Choose front-end framework tool

3. Choose web-hosting service
4. Choose data source

8.4. DEVELOP DATABASE (03-08-2019)

1. Set up web server
2. Write up database generation code
3. Populate database with data

8.5. FINISH DATABASE AND PRESENTATION (03-18-2019)

1. Develop website to access database
2. Make poster for presentation
3. Present final project to class

9. WORK DISTRIBUTION

With regards to selecting a database idea, our group discussed what level of interest we had in several different ideas to provide input on entity types and selection. During the brainstorming phase, we all provided ideas and input regarding what entities should exist inside an NBA database. We discussed at length regarding what relationships and references should exist between the entities that we identified. When it comes to implementing the database, Jeremy used his experience with MySQL to write code for the database and developed the ER Diagram and RM Schema on draw.io. When it comes to creating the documents for the project, Aaron was able to compile team ideas and record them with proper formatting. With regards to ensuring group collaboration, Brandon provided the framework for pushing group documents to Github and helped practice version control. Nathan worked with the group on setting up the web hosting service for the project. All members of Team Avatar will contribute equally to the voter database application. It's anticipated that each team member will produce 25% of the work overall; however, discrepancies within individual deliverables are expected.

10. TOOLING ASSESSMENT

In this section, we will outline the tools we chose to use for this project and its reasons behind the decision.

10.1. DBMS: MYSQL WORKBENCH CE

Team Avatar has chosen this DBMS because of its ease of use, popularity, and the fact that one of our members has experience with MySQL Workbench CE.

10.2. UI: HTML/CSS

For the front-end side, we have chosen a simple html page with buttons and multiple pages for each table would be sufficient enough for this project as the main focus is the database, not the UX.

10.3. VERSION CONTROL: GITHUB

GitHub is a very well-known platform for collaborators to share their code and it's free to use. Besides that, most of our members have prior experience in using GitHub as a version control platform.

10.4. WEB HOSTING SERVICE: AWS OR USBWEBSERVER

For now, our team is still exploring the possibilities of two web hosting services; Amazon Web Service and USBWebServer. Both are free, and our final decision will depend on ease of use.