**Part I: Requirements Gathering & E-R Diagram**

***Deliverable 1: Executive Summary***

This database is built to store detailed MLB player statistics from the last 20 years, along with information about teams, divisions, and leagues. The main data includes individual player stats like home runs, RBIs, and strikeouts, as well as personal info like name, birthdate, position, and nationality. It also stores team rosters for each season and shows which teams belong to which divisions and leagues. By linking players to their teams and tracking their performance over time, the database can help users analyze historical data in depth.

The main users of this database would include data analysts, sports journalists, team managers, and baseball fans. They will use it to spot trends in player performance, compare stats between players, track roster changes, and find out which players or teams led certain categories in different seasons.

Some examples of questions this database can answer include:

* Who hit the most home runs in 2020?
* How has a player's batting average changed over the past five seasons?
* What was the New York Yankees' roster in 2015?
* Which American League team had the most wins in 2019?

The ER diagram helps answer these questions by organizing the data across several tables and showing how they are related. The Stats table stores performance data by season, while the Roster table links players to their teams each year. The Team and Division tables make it easy to group data for teams and leagues. This structure makes the database a powerful tool for analyzing player stats and team performance over time.

***Deliverable 2: E-R Diagram***

A screenshot of a computer

Description automatically generated

**Part II: Database Schema**

***Deliverable 1: CREATE TABLE Statements***

CREATE TABLE Division (

divisionID INT PRIMARY KEY,

leagueName VARCHAR(50) NOT NULL,

divisionLocation VARCHAR(50) NOT NULL

);

CREATE TABLE Team (

teamID INT PRIMARY KEY,

cityName VARCHAR(100) NOT NULL,

mascotName VARCHAR(50) NOT NULL,

divisionID INT,

FOREIGN KEY (divisionID) REFERENCES Division(divisionID)

);

CREATE TABLE Player (

playerID INT PRIMARY KEY,

firstName VARCHAR(50) NOT NULL,

lastName VARCHAR(50) NOT NULL,

birthDate DATE NOT NULL,

originCountry VARCHAR(50),

height INT,

weight INT,

position VARCHAR(50),

currentTeamID INT,

FOREIGN KEY (currentTeamID) REFERENCES Team(teamID)

);

CREATE TABLE Roster (

rosterID INT PRIMARY KEY,

teamID INT,

playerID INT,

season INT NOT NULL,

FOREIGN KEY (teamID) REFERENCES Team(teamID),

FOREIGN KEY (playerID) REFERENCES Player(playerID)

);

CREATE TABLE Stats (

statsID INT PRIMARY KEY,

playerID INT,

season INT NOT NULL,

gamesPlayed INT,

hits INT,

runs INT,

homeRuns INT,

RBI INT,

battingAverage FLOAT,

wins INT,

strikeOuts INT,

ERA FLOAT,

WHIP FLOAT,

saves INT,

FOREIGN KEY (playerID) REFERENCES Player(playerID)

);

***Deliverable 2: SELECT Statements***

Who hit the most home runs in 2020?

SELECT P.firstName, P.lastName, S.homeRuns

FROM Stats S

JOIN Player P ON S.playerID = P.playerID

WHERE S.season = 2020

ORDER BY S.homeRuns DESC

LIMIT 1;

How has a player's batting average changed over the past five seasons?

SELECT S.season, S.battingAverage

FROM Stats S

JOIN Player P ON S.playerID = P.playerID

WHERE P.firstName = 'Mike' AND P.lastName = 'Trout'

ORDER BY S.season DESC;

What was the New York Yankees' roster in 2015?

SELECT P.firstName, P.lastName, P.position

FROM Roster R

JOIN Player P ON R.playerID = P.playerID

JOIN Team T ON R.teamID = T.teamID

WHERE T.mascotName = 'Yankees' AND R.season = 2015;

Which American League team had the most wins in 2019?

SELECT T.cityName, T.mascotName, SUM(S.wins) AS totalWins

FROM Stats S

JOIN Player P ON S.playerID = P.playerID

JOIN Roster R ON P.playerID = R.playerID

JOIN Team T ON R.teamID = T.teamID

JOIN Division D ON T.divisionID = D.divisionID

WHERE S.season = 2019

AND D.leagueName = 'American League'

GROUP BY T.cityName, T.mascotName

ORDER BY totalWins DESC

LIMIT 1;