Tristan Brisker

Ricardo Figueroa

Team 13

**Project Title: NFL Database**

Project Objective: The objective of this project is to create a database of the 32 NFL teams and their associated players. The expected outcomes include detailed information about the teams, players and related statistics. The database will be designed to facilitate efficient querying, analysis and reporting.

**Scope of the Project:**

The scope of this project includes the following aspects of the NFL operations:

* NFL players with their statistics, positions, and years in league
* Teams with their salary caps, owners, general manager, end of season record, division, and conference
* Coaches separated into their respective positional groups (Offensive Coordinator, Defensive Coordinator, Special Teams Coordinator), what team they belong to, and years of experience coaching
* Regular Season Statistics includes record, ranks, points for, and points against
* Stadium that provides the stadium name, mascot, city, state, and seating capacity

**ER Diagram:**

Below is the ER diagram of the NFL Database.

A diagram of a game

AI-generated content may be incorrect.

**Relational Model:**

Based on the ER diagram the following schemas were created.

1. The entity sets and their attributes below, with primary keys underlined.

player (player\_id, name, position, jersey\_number, team\_id)

position (Position\_id, Position\_name)

coach (coach\_id, name, coach\_type, team\_name, age, experience\_years, team\_id)

team (team\_id, season\_rank, team\_name, city, division, head\_coach\_id)

game (game\_id, home\_team\_id, away\_team\_id, game\_date, stadium\_id)

stadium (stadium\_id, name, city, capacity, team\_id)

1. Schema derived from relationship sets in the E-R diagram

player position (position, position\_name, position\_id)

coaches (coach\_id, team\_id, team\_name)

plays\_for (player\_id, team\_id, team\_name)

leads (coach\_id, head\_coach\_id)

plays\_in (team\_id, home\_team\_id, away\_team\_id)

is\_played\_in (stadium\_id, team\_id)

To overcome the redundancy and follow the normalization principle we reduced the schema to the following final schemas. Most of the relations between the tables have been condensed into the attributes that the tables share (foreign keys). The only table that has a relationship still in use is the player position given that the position table is purely dependent on the player table.

Player (player\_id, name, position, jersey\_number, team\_id)

Position (Position\_id, Position\_name)

Coach (coach\_id, name, coach\_type, team\_name, age, experience\_years, team\_id)

Team (team\_id, season\_rank, team\_name, city, division, head\_coach\_id)

Game (game\_id, home\_team\_id, away\_team\_id, game\_date, stadium\_id)

Stadium (stadium\_id, name, city, capacity, team\_id)

**Schema Diagram for NFL Database:**

Based on the final schema above we draw our schema diagram for our database, along with primary key and foreign key constraints as displayed below. Each relation appears as a box with the relation name at the top in bold and the attributes listed below. Primary key attributes are shown underlined. Foreign key constraints appear as arrows from the foreign key attributes of the referencing relation to the key of the referenced relation.



**Dataset Description:**

* Team – Each team is identified by their name. The ‘team’ table is the most important holding the team conference, team performance and links the ‘players’ and ‘coaches’ tables.
* Players- This table contains information on individual players, including current team, and link to ‘regular season stats.
* Coaches- This table tracks the NFL coaches, including the Head Coach and coordinators. It will also hold how many years they have been a coach.
* Regular Season Stats- This table records player performance for each player and team for the season.
* Stadium- This table records the stadium name, capacity, city and mascot for all the teams, and links to the ‘team’ table.

**Table Creation:**

Details of each table inserted are attached below:



**Data Population:**

The data was collected from the NFL website and the insert queries were generated using ChatGPT. After using the LLM, we searched through the information to verify that all lines were correctly populated. An example of what we provided to ChatGPT was the initial structure of the table (ex. INSERT INTO Coach (name, coach\_type, team\_name, age, experience\_years, team\_id)). With this, it would generate the data required using the website.

**Query Questions:**

Here are some questions that can be answered using the data:

1. Which teams are in the NFC East, what are the stadium names, stadium capacity and total number of players on each team?
2. What are the names, positions, and experience levels of coaches on the New England Patriots?
3. How many players are on the roster for each team, ranked within their division?
4. Which head coaches lead teams that play in stadiums with more than 70,000 capacity?
5. What games are scheduled on a given date, including home & away teams, and days until the match?
6. What are the average years of experience from coaches from each division?
7. Which teams have more than the lowest number of players per team?
8. What is the percentage of players in each division compared to the total number of players?
9. Which teams have coaches that are higher than the average age of the entire league?
10. Which players in the league have the same name as any coach in the league?

**Data Source:**

https://www.nfl.com