

PRIMERA Y SEGUNDA ENTREGA BASES DE DATOS

Liga de fútbol

Nombres:

- Juan Pablo Simbaqueba
- Miguel Salamanca
- Camila Patarroyo

1. Buscar y seleccionar una o varias fuentes de datos reales. Por ejemplo, tráfico, servicios públicos, accidentalidad, movilidad, salud, gobierno, etc. Se pueden considerar datos propios de algún proyecto o los datos publicados en proyectos de datos abiertos como <https://www.datos.gov.co/> , cuyos datos deben haberse publicado a partir del segundo semestre de 2022. En el segundo caso, es importante incluir la URL del proyecto y de los datos seleccionados.

La fuente que usaremos para extraer los datos es <https://www.transfermarkt.co/>. El cual es una página web que tiene información del mundo del futbol, como puntuaciones, resultados, noticias de transferencia y calendarios de liga, nosotros lo hicimos con la liga de futbol de Colombia.

2. Con base en los datos seleccionados, definir un problema de datos identificando el mayor número posible de reglas de negocio.

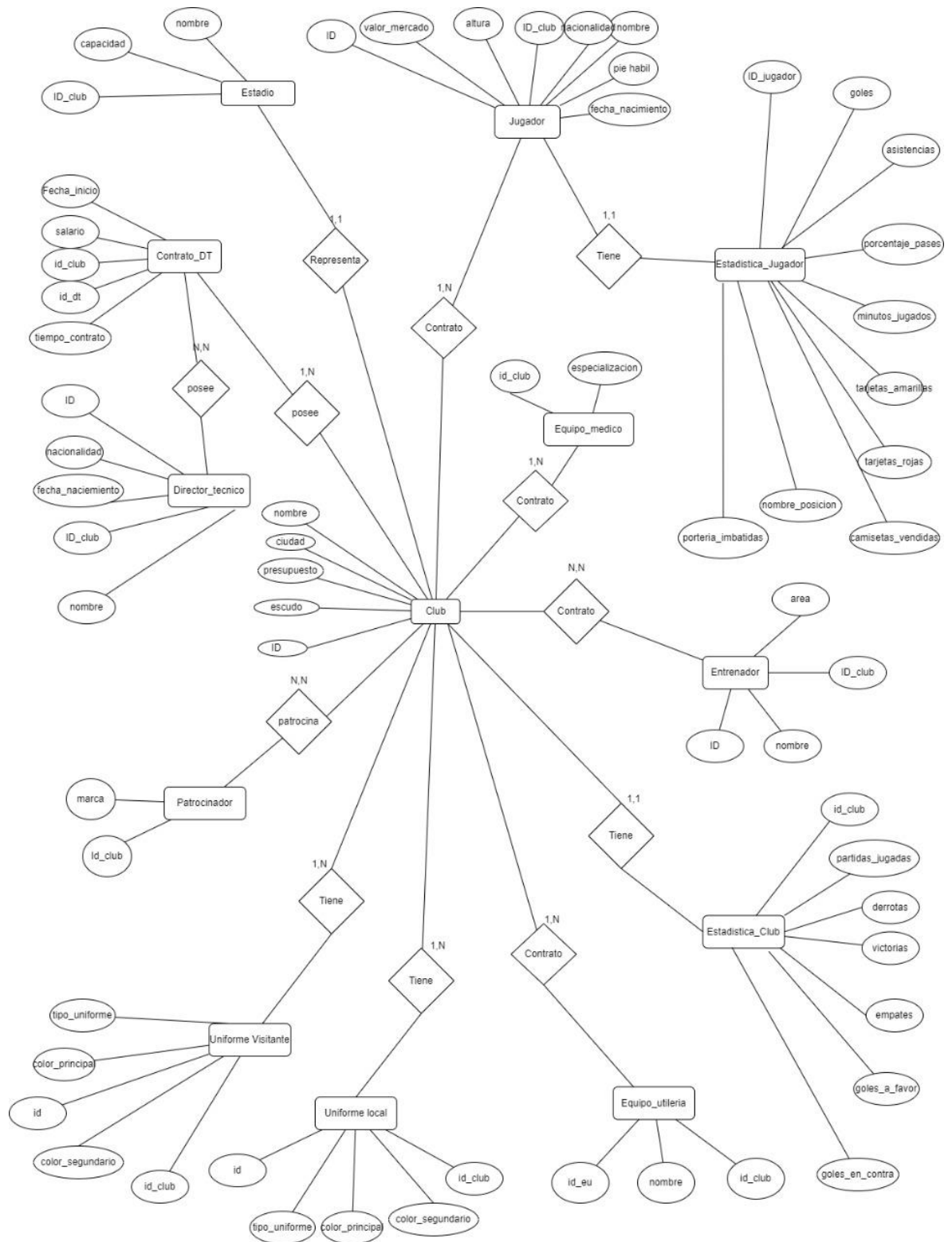
Esta propuesta busca facilitar la búsqueda de información por parte de los usuarios que necesiten o quieran por su propia cuenta, obtener datos acerca de su liga de preferencia. Hoy en día hay muchos usuarios dentro del mundo del futbol que quisieran obtener una gran cantidad de datos para realizar un buen análisis sobre su equipo o sobre algún jugador, ya sea solo para adquirir conocimiento o para algún otro tema que incluya un objetivo monetario para esta persona, como lo es el mundo de las apuestas, o también si se necesitan sacar datos para los mismos funcionarios del club para manejar el tema de las finanzas e igualmente para los funcionarios de las propias ligas. En cuanto al mundo de las apuestas, existen muchas personas que basan su economía en esta actividad, ya sea por parte de las casas de apuestas como también de sus usuarios, el hecho de tener una base de datos que contenga cada estadística del club y sus jugadores permitirá una mejor predicción y facilitara el hecho de sacar un indicador de multiplicación para las apuestas de cada partido. Como segunda problemática a superar, actualmente cada equipo de futbol mueve mucho dinero fuera de sus partidos, la realización de esta propuesta busca facilitar para los usuarios la búsqueda de cualquier producto que quieran adquirir que esté relacionado con el club, ya sean sus camisetas, sus sudaderas, sus equitaciones, o también

las boletas de los partidos, sin tener que salir del lugar donde encuentran las estadísticas y resultados del club. Todo al alcance de un clic desde la comodidad ya sea de su celular, computador o tableta.

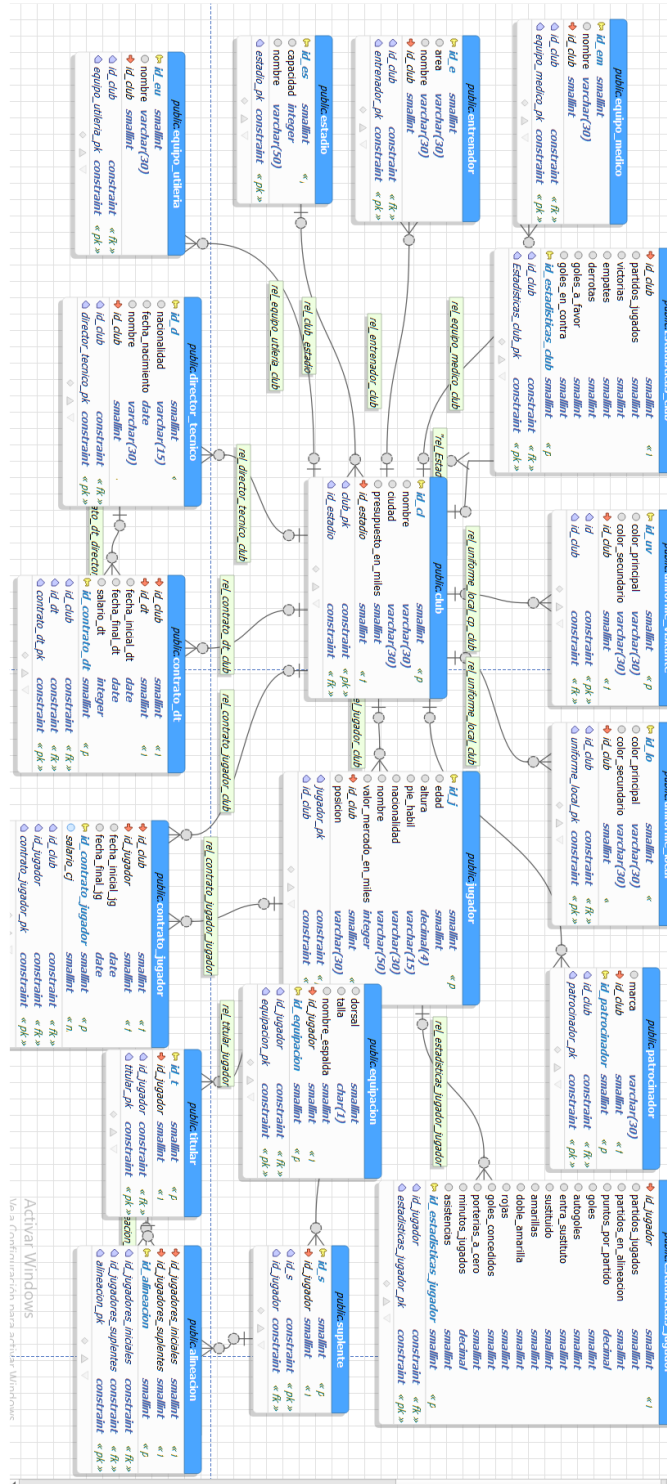
3. Identificar y describir las entidades, atributos y relaciones de las fuentes de datos.

La base de datos cuenta con 13 tablas las cuales son: club, contrato dt, director técnico, entrenador, equipo médico, equipo utilería, estadio, estadísticas club, estadísticas jugador, jugador, patrocinador, uniforme local y uniforme visitante.

4. Construir el diagrama Entidad Relación con base en las reglas que se definieron a partir de las fuentes de datos seleccionadas. Utilizar notación CHEN.



5. Construir el diagrama relacional basado en el diagrama ER. Utilizar notación CHEN y/o pgModeler.



1. Construir el diagrama relacional normalizado en tercera forma normal.

Para la edición del plano inicial de las tablas tuvimos en cuenta la normalización 3NF que se estructura de la siguiente forma:

Reglas para 1NF 1.

Cada registro debe ser único, sin filas duplicadas 2. Cada celda debe contener un valor, debe verificar que no existan atributos multivalor Todo nuestro modelo ya estaba normalizado en 1NF, por lo cual no tuvimos que hacer ninguna modificación en este aspecto, ya que cumplía las condiciones.

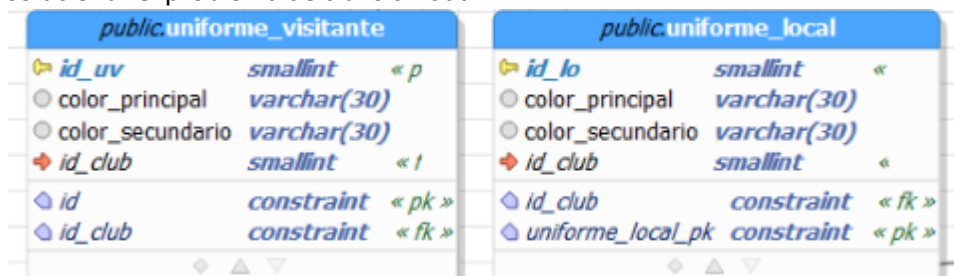
Regla para 2NF 1.

Debe satisfacer 1NF 2. Si la clave principal (PK) de una tabla consta de uno y solo una columna, automáticamente satisface 2NF 3. Si hay una clave principal compuesta, cada columna no clave debe depender de toda la llave. Se deben eliminar dependencias parciales de la llave. Existían unas tablas en nuestro modelo que no tenían llave primaria como la tabla estadísticas club, lo cual resolvimos creando el atributo id estadísticas club porque los demás atributos no eran los apropiados para ser llave primaria de esta tabla.



Reglas para 3NF 1.

Satisfacer 2NF 4 2. Sin dependencias transitivas: Las columnas no claves (que no forman parte de la PK), no pueden depender de otras columnas que tampoco sean no clave. En nuestro caso en específico inicialmente teníamos una tabla llamada uniforme con los siguientes atributos: id club, tipo de uniforme, color primario, color secundario Note que hay una dependencia transitiva entre los colores, el tipo y el id del club, ya que un club tiene dos tipos de uniformes y varios colores pueden ser de muchos clubs y tipos de uniforme. Por lo que decidimos crear dos tablas, uniforme visitante y uniforme local el cual solucionar el problema de transitividad.



El modelo se puede ver en el punto 5 de la entrega 1 ya que la hicimos teniendo en cuenta la normalización.

2. Implementar en PostgreSQL la base de datos normalizada en tercera forma normal.

Archivo txt adjuntado

3. Cargar todos los datos de las fuentes de datos en la base de datos previamente creada.

Archivos csv adjuntados

4. Identificar y describir al menos cuatro posibles escenarios de análisis que podrían realizarse con los datos cargados en la base de datos.

Con la base de datos que tenemos podríamos hacer:

- Un ranking de los mejores goleadores del torneo
- Buscar los colores característicos de cada club
- Presupuesto disponible para cada uno de los directores técnicos
- Jugadores más influyentes

5. Desarrollar en python los módulos necesarios para conectarse a la base de datos y consultar los datos de todas las tablas.

Archivo py adjuntado

6. Crear un repositorio GIT en github o gitlab con los avances del proyecto. Todos los miembros del grupo deben hacer commit al repositorio, como evidencia de su participación en el desarrollo del proyecto.

https://github.com/Camila-pv/Liga_Colombiana_Futbol.git