

Base de Datos I

Trabajo Práctico Especial

2^{do} Cuatrimestre 2024

1. Objetivo

El objetivo de este Trabajo Práctico Especial es aplicar los conceptos de SQL Avanzado (PSM, Triggers) vistos a lo largo del curso, para implementar funcionalidades y restricciones no disponibles de forma estándar (que no pueden resolverse con Primary Keys, Foreign Keys, etc.).

2. Modalidad

El Trabajo Práctico estará disponible en el Campus a partir del 14/11/2024, indicándose allí mismo la fecha de entrega.

Se incluye junto con el enunciado el archivo: **jugadores-2022.csv**.

El TP deberá realizarse en grupos de 4 alumnos y entregarse a través de la plataforma Campus ITBA hasta la fecha allí indicada.

3. Descripción del Trabajo

Se tiene un esquema de base de datos con jugadores de equipos de primera división en determinada temporada, donde cada equipo asigna dorsales únicos que identifican a cada jugador y cada jugador juega para un equipo determinado.

Viendo este ejemplo resumido:

Equipo	Dorsal	Jugador
Independiente	8	Lucas González
Boca	9	Miguel Merentiel
River	8	Rodrigo Aliendro
Independiente	9	Santiago Hidalgo

Las dependencias funcionales mencionadas se pueden expresar como:

- **Equipo Dorsal → Jugador**
- **Jugador → Equipo**

Las claves candidatas de este esquema son:

- **Dorsal + Equipo**
- **Dorsal + Jugador**

Como se desea normalizar a BCNF, rápidamente se ve que Jugador → Equipo viola BCNF ya que Jugador no es clave.

Entonces se genera un nuevo subesquema **R1(Jugador, Equipo)** con **F1 = { Jugador → Equipo }**, clave **Jugador**.

El esquema restante al eliminar el atributo Equipo es **R2(Jugador, Dorsal)** con **F2 = { }**, clave **Jugador + Dorsal**.

Como $F1 \cup F2 = \{ \text{Jugador} \rightarrow \text{Equipo} \}$ y $(\text{Equipo}, \text{Dorsal})^+ = \{ \text{Equipo}, \text{Dorsal} \}$, entonces **se pierde la dependencia Equipo Dorsal → Jugador**

Notar que al descomponer la instancia de ejemplo quedarían R1 y R2 de la siguiente manera:

R1	
Jugador	Equipo
Lucas González	Independiente
Miguel Merentiel	Boca
Rodrigo Aliandro	River
Santiago Hidalgo	Independiente

R2	
Jugador	Dorsal
Lucas González	8
Miguel Merentiel	9
Rodrigo Aliandro	8
Santiago Hidalgo	9

Nada impediría insertar en R1 lo siguiente, ya que **Jugador → Equipo** sigue valiendo

R1	
Jugador	Equipo
Sergio Fabián Ortiz	Independiente

y luego insertar en R2

R2	
Jugador	Dorsal
Sergio Fabián Ortiz	8

Con estas dos instancias y juntándolas por el atributo en común Jugador, obtendríamos:

Equipo	Dorsal	Jugador
Independiente	8	Lucas González
Boca	9	Miguel Merentiel

River	8	Rodrigo Aliendro
Independiente	9	Santiago Hidalgo
Independiente	8	Sergio Fabián Ortiz

Y claramente esta instancia viola **Equipo Dorsal → Jugador**

Como igualmente se quiere mantener esta descomposición BCNF y no pasar a 3FN, entonces debemos implementar los PSMs / Triggers necesarios para asegurar que la dependencia perdida se siga cumpliendo.

En este Trabajo Práctico se implementará este procedimiento para 2 tablas que serán nuestras R1 (con más campos) y R2 respectivamente. La tabla **futbolista (nombre, posicion, edad, altura, pie, fichado, equipo_anterior, valor_mercado, equipo)** y la tabla **dorsal (jugador, dorsal)**. Ambas deberán ser creadas de acuerdo a los requerimientos y los datos provistos. Es por ello, que se proveerá un archivo CSV (Comma Separated Values) con información para realizar el trabajo.

El archivo es una reducción de un dataset que tomamos del sitio de Kaggle, llamado “*Base de Datos Fútbol Argentino (2008-2022)*”. En esta ocasión vamos a utilizar sólo datos de la temporada 2022. Este dataset reducido contiene información de jugadores y equipos de primera división, el cual permite hacer un análisis de estadísticas, entre otros.

Las columnas del archivo son:

- **Jugador:** Nombre del jugador en cuestión
- **Posición:** Posición en la que juega en el equipo actual (portero, defensa, etc.)
- **Edad:** Edad actual del jugador en cuestión
- **Altura:** Altura del jugador en cuestión expresada en metros
- **Pie:** Pie con el que patea mejor el jugador en cuestión
- **Fecha del fichaje:** Fecha en la que el jugador fue fichado en el equipo actual
- **Equipo anterior:** Nombre del equipo anterior en el que jugó el jugador en cuestión
- **Valor del mercado:** Valor del mercado del jugador al 2022 expresado en dólares
- **Equipo:** Nombre del equipo actual donde juega jugador en cuestión

Como en el archivo de datos csv no viene el dato del dorsal asignado, es necesario generarlo automáticamente durante de la carga de datos, teniendo en cuenta los usos y costumbres del fútbol argentino y considerando la posición del jugador. Es decir, el arquero lleva el 1 generalmente, sino el 12; los centrales pueden llevar el 2 o el 6; los laterales pueden llevar el 3 o el 4, etc. Y luego si esos números "principales" no están libres, tomar el primero disponible del resto del equipo.

Si seguimos la lógica, la asignación sería de la siguiente manera:

- Portero = 1 o 12
- Defensa, Defensa central = 2 o 6
- Lateral izquierdo = 3
- Lateral derecho = 4
- Pivote = 5
- Mediocentro, Centrocampista, Interior derecho, Interior izquierdo = 8
- Mediocentro ofensivo, Mediapunta = 10
- Extremo derecho, Extremo izquierdo = 7 o 11 respectivamente
- Delantero, Delantero centro = 9

Si ya está asignado el dorsal, entonces podemos asignar el primero libre a partir del 13 para el equipo en cuestión.

Por último, el administrador de la base de datos desea realizar periódicamente un análisis de los jugadores y los equipos para recomendar posibles acciones a los distintos equipos y jugadores.

En resumen, la finalidad de este Trabajo Práctico Especial consiste en implementar lo antes descrito. Específicamente se debe hacer lo siguiente:

- Crear las 2 tablas *futbolista* y *dorsal*
- Implementar un trigger para garantizar la dependencia funcional perdida y una función para asignar el dorsal más adecuado de cada jugador durante la carga de los mismos
- Importar los datos y cargar las tablas *futbolista* y *dorsal*
- Implementar la función para analizar las estadísticas de los equipos y jugadores

4. Explicación paso a paso

a) Creación de las tablas *futbolista* y *dorsal*

Debe crearse la tabla *futbolista* con los tipos de datos adecuados para almacenar los datos procedentes del archivo CSV. Definir la clave y las restricciones según corresponda.

Debe crearse la tabla *dorsal* con los tipos de datos adecuados para almacenar los datos del nombre del jugador y del número del dorsal a asignar. Definir la clave y las restricciones según corresponda.

b) Implementación de un trigger para garantizar la dependencia funcional perdida y de una función para asignar el dorsal más adecuado

Debe implementarse mediante un trigger, el chequeo que permita garantizar que la dependencia **Equipo Dorsal → Jugador** no se pierde.

Por ejemplo:

- Si el usuario realiza la operación *insert into futbolista values ('Lucas González', 'Mediocentro', 21, 1.69, 'derecho', '01/08/2020', 'CA Independiente II', 1700000, 'Independiente');* y si asumimos que las tablas comienzan vacías, se asigna el dorsal **8** contemplando la columna *posición*, quedando las tablas *futbolista* y *dorsal* de la siguiente manera:

FUTBOLISTA								
nombre	posicion	edad	altura	pie	fichado	equipo_anterior	valor_mercado	equipo
Lucas González	Mediocentro	21	1.69	derecho	01/08/2020	CA Independiente II	1.700.000	Independiente

DORSAL	
jugador	dorsal
Lucas González	8

• Si luego el usuario realiza la operación *insert into futbolista values ('Miguel Merentiel, 'Delantero centro', 35, 1.76, 'derecho', '02/02/2023', 'SE Palmeiras', 2500000, 'Boca Juniors');* se asigna el dorsal **9** contemplando la columna *posición*, quedando las tablas futbolista y dorsal de la siguiente manera:

FUTBOLISTA								
nombre	posicion	edad	altura	pie	fichado	equipo_anterior	valor_mercado	equipo
Lucas González	Mediocentro	21	1.69	derecho	01/08/2020	CA Independiente II	1.700.000	Independiente
Miguel Merentiel	Delantero centro	35	1.76	derecho	02/02/2023	SE Palmeiras	2.500.000	Boca Juniors

DORSAL	
jugador	Dorsal
Lucas González	8
Miguel Merentiel	9

• Si luego el usuario realiza la operación *insert into futbolista values ('Sergio Fabián Ortiz, 'Mediocentro', 23, 1.74, 'derecho', '01/01/2023', 'CA Independiente II', null, 'Independiente');* se asigna el dorsal más adecuado, en este caso el **13** contemplando la columna *posición* y porque el 8 ya está ocupado por Lucas González en el equipo de Independiente, quedando las tablas futbolista y dorsal de la siguiente manera:

FUTBOLISTA								
nombre	posicion	edad	altura	Pie	fichado	equipo_anterior	valor_mercado	equipo
Lucas González	Mediocentro	21	1.69	derecho	01/08/2020	CA Independiente II	1.700.000	Independiente
Miguel Merentiel	Delantero centro	35	1.76	derecho	02/02/2023	SE Palmeiras	2.500.000	Boca Juniors
Sergio Fabián Ortiz	Mediocentro	23	1.74	Izquierdo	01/01/2023	CA Independiente II	null	Independiente

DORSAL	
jugador	dorsal
Lucas González	8
Miguel Merentiel	9
Sergio Fabián Ortiz	13

También es importante considerar el manejo de errores para evitar el ingreso de datos incorrectos. Puede lanzar mensajes, excepciones, lo que crea conveniente, siempre que se entienda la razón de la misma. Por ejemplo, si el usuario quiere insertar un jugador ya cargado en la tabla de futbolista.

c) Importación de los datos

Utilizando el comando COPY de PostgreSQL, se deben importar TODOS los datos del archivo csv en las tablas creadas en el punto a). El archivo csv provisto por la cátedra NO puede ser modificado.

d) Análisis de equipos y jugadores

El administrador de la base de datos desea realizar periódicamente un análisis de los jugadores y los equipos para recomendar posibles acciones a los distintos equipos y jugadores.

Se pide crear la función *analisis_jugadores(dia)* que recibe como parámetro una fecha, que se tomará como fecha base a partir de la cual se considerarán solo los jugadores que hayan sido fichados después de dicha fecha, la cual genere un reporte mostrando un resumen de las estadísticas de:

- Los pies (derecho e izquierdo) por mes de fichaje, mostrando la cantidad de jugadores involucrados, el promedio de edad, el promedio de altura y el valor máximo de jugador del mercado, ordenado de menor a mayor por pie y mes
- Los equipos, mostrando la fecha mínima de fichaje dentro de ese equipo, la cantidad de jugadores involucrados, el promedio de edad, el promedio de altura y el valor máximo de jugador del mercado, ordenado de mayor a menor valor de mercado
- Los dorsales principales, mostrando la fecha mínima de fichaje dentro de ese equipo, la cantidad de jugadores involucrados, el promedio de edad, el promedio de altura y el valor máximo de jugador del mercado, ordenado de mayor a menor valor de mercado

El reporte tendrá las siguientes características:

- I. Título del reporte: "ANALISIS DE JUGADORES Y EQUIPOS"
- II. Encabezado de columnas:
"Variable Fecha Qty Prom_Edad Prom_Alt Valor #"
- III. Por cada Variable tiene que aparecer un renglón en el reporte, con la información pedida para cada punto, adicionando el número de línea # (la cual se reinicia en 1 al cambiar el valor de la variable en el caso de los pies)

En caso de que no existieran datos para fechas mayores a la fecha ingresada, no se debe mostrar nada (ni siquiera el encabezado del reporte).

La función debe manejar posibles errores.

Por ejemplo:

- Si invocamos `analisis_jugadores('22/07/2022')` se obtiene un informe que comienza de la siguiente manera (considerando los datos ya ingresados del archivo CSV):

```
INFO: -----
INFO: -----ANALISIS DE JUGADORES Y EQUIPOS -----
INFO: -----
INFO: -----
INFO: Variable-----Fecha-----Qty--Prom_Edad--Prom_Alt--Valor-----#---
INFO: -----
INFO: Pie: derecho..... 2022-07      3    29.3    1.84    800000    1
INFO: Pie: derecho..... 2022-08      4    22.5    1.82    2800000    2
INFO: Pie: derecho..... 2023-01     15    23.4    1.80    4000000    3
INFO: Pie: derecho..... 2023-02      2    29.0    1.80    2500000    4
INFO: Pie: izquierdo..... 2022-07      1    22.0    1.92    450000    1
INFO: Pie: izquierdo..... 2022-08      1    17.0    1.76    1500000    2
INFO: Pie: izquierdo..... 2023-01      5    26.2    1.79    6500000    3
INFO: Pie: izquierdo..... 2023-07      1    24.0    1.72    1200000    4
INFO: -----
INFO: River Plate..... 2023-01-01      3    23.0    1.82    6500000    1
INFO: Boca..... 2022-08-08      8    27.5    1.82    4000000    2
INFO: Newell's Old Boys..... 2022-08-05      2    19.5    1.83    2800000    3
INFO: San Lorenzo..... 2022-07-26      3    28.3    1.85    2300000    4
INFO: Rosario Central..... 2022-08-08      1    17.0    1.76    1500000    5
INFO: Argentinos Juniors..... 2022-08-12      1    22.0    1.70    700000    6
INFO: Estudiantes de La Plata..... 2022-07-28      3    22.7    1.83    700000    7
INFO: Racing..... 2023-01-03      1    21.0    1.71    675000    8
INFO: Tigre..... 2023-01-01      3    23.0    1.77    475000    9
INFO: Huracan..... 2022-07-22      4    26.5    1.78    450000    10
INFO: Godoy Cruz Antonio Tomba..... 2023-01-01      1    22.0    1.85    100000    11
INFO: Gimnasia y Esgrima La Plata... 2023-01-01      1    21.0    1.73    50000    12
INFO: Velez Sarsfield..... 2023-01-01      1    22.0    1.74    25000    13
INFO: Arsenal..... 2023-01-01      1    21.0    1.91    25000    14
INFO: -----
INFO: Dorsal: 6..... 2023-01-01      1    21.0    1.85    6500000    1
INFO: Dorsal: 2..... 2023-01-16      2    25.0    1.82    4000000    2
INFO: Dorsal: 1..... 2022-08-05      1    19.0    1.87    2800000    3
INFO: Dorsal: 3..... 2022-08-08      1    17.0    1.76    1500000    4
INFO: Dorsal: 12..... 2022-08-08      1    22.0    1.92    700000    5
INFO: Dorsal: 11..... 2023-01-03      1    20.0    1.79    475000    6
INFO: Dorsal: 4..... 2023-01-01      1    28.0    1.75    400000    7
INFO: -----
```

- Si invocamos `analisis_jugadores('22/01/2022')` se obtiene un informe que comienza de la siguiente manera (considerando los datos ya ingresados del archivo CSV):

```

INFO: -----ANALISIS DE JUGADORES Y EQUIPOS -----
INFO: -----
INFO: -----
INFO: Variable-----Fecha-----Qty--Prom_Edad--Prom_Alt--Valor-----#--
INFO: -----
INFO: Pie: derecho..... 2022-01      14      23.5      1.80      10000000      1
INFO: Pie: derecho..... 2022-02      26      25.3      1.78      23000000      2
INFO: Pie: derecho..... 2022-03       2      26.5      1.77      475000      3
INFO: Pie: derecho..... 2022-05       2      21.0      1.79      350000      4
INFO: Pie: derecho..... 2022-06      16      25.1      1.79      4000000      5
INFO: Pie: derecho..... 2022-07      50      25.8      1.80      4000000      6
INFO: Pie: derecho..... 2022-08       4      22.5      1.82      2800000      7
INFO: Pie: derecho..... 2023-01      15      23.4      1.80      4000000      8
INFO: Pie: derecho..... 2023-02       2      29.0      1.80      2500000      9
INFO: Pie: izquierdo..... 2022-01       8      23.0      1.79      3200000      1
INFO: Pie: izquierdo..... 2022-02      10      24.5      1.75      5000000      2
INFO: Pie: izquierdo..... 2022-06       7      23.0      1.80      1500000      3
INFO: Pie: izquierdo..... 2022-07      17      24.1      1.80      1000000      4
INFO: Pie: izquierdo..... 2022-08       1      17.0      1.76      1500000      5
INFO: Pie: izquierdo..... 2023-01       5      26.2      1.79      6500000      6
INFO: Pie: izquierdo..... 2023-07       1      24.0      1.72      1200000      7
INFO: -----
INFO: River Plate..... 2022-01-27      11      22.5      1.76      23000000      1
INFO: Velez Sarsfield..... 2022-01-25      11      24.6      1.78      4500000      2
INFO: Boca..... 2022-01-24      13      26.2      1.81      4000000      3
INFO: Estudiantes de La Plata..... 2022-01-26      13      23.5      1.79      3800000      4
INFO: San Lorenzo..... 2022-02-11      10      24.7      1.83      3000000      5
INFO: Racing..... 2022-01-26       6      25.5      1.76      3000000      6
INFO: Newell's Old Boys..... 2022-01-26      11      24.5      1.80      2800000      7
INFO: Independiente..... 2022-01-28      10      24.3      1.77      2700000      8
INFO: Rosario Central..... 2022-02-03      14      26.2      1.79      2500000      9
INFO: Gimnasia y Esgrima La Plata... 2022-02-02       6      26.0      1.78      2500000      10
INFO: Huracan..... 2022-01-25      14      24.2      1.78      2400000      11
INFO: Argentinos Juniors..... 2022-01-25       7      25.3      1.73      2000000      12
INFO: Colon..... 2022-01-24      12      25.1      1.81      1600000      13
INFO: Tigre..... 2022-02-08       7      23.4      1.81      1500000      14
INFO: Arsenal..... 2022-01-22      13      26.1      1.81      1300000      15
INFO: Lanus..... 2022-01-22      10      24.7      1.81      1300000      16
INFO: Godoy Cruz Antonio Tomba..... 2022-01-22      14      23.1      1.81      800000      17
INFO: -----
INFO: Dorsal: 9..... 2022-01-26       5      26.2      1.79      23000000      1
INFO: Dorsal: 7..... 2022-01-29       3      22.0      1.73      10000000      2
INFO: Dorsal: 6..... 2022-01-24       5      24.8      1.85      6500000      3
INFO: Dorsal: 5..... 2022-02-10       3      23.3      1.79      4500000      4
INFO: Dorsal: 2..... 2022-02-10       5      26.2      1.85      4000000      5
INFO: Dorsal: 4..... 2022-02-07       3      23.7      1.76      3200000      6
INFO: Dorsal: 3..... 2022-01-25       5      24.2      1.74      3200000      7
INFO: Dorsal: 10..... 2022-02-01       6      27.5      1.74      3000000      8
INFO: Dorsal: 11..... 2022-02-10       6      22.3      1.72      3000000      9
INFO: Dorsal: 1..... 2022-01-31       4      22.0      1.89      2800000      10
INFO: Dorsal: 8..... 2022-07-09       1      20.0      1.77      2800000      11
INFO: Dorsal: 12..... 2022-01-22       4      23.0      1.87      1800000      12
INFO: -----

```

- Si invocamos `analisis_jugadores('22/07/2023')` no se obtiene nada

4. Entregables

Los alumnos deberán entregar los siguientes documentos:

- El script sql **funciones.sql** con el código necesario para crear las tablas, las funciones y los triggers
- Un informe que debe contener:
 - El rol de cada uno de los participantes del grupo. Si bien en el TP deben estar involucrados todos los integrantes, se debe asignar un rol de supervisión de cada una de las tareas. Mínimamente los roles son: encargado del informe, encargado de las funciones, encargado del trigger, encargado del funcionamiento global del proyecto y encargado de investigación. Pueden asignarse más roles en caso de requerirse
 - Todo lo investigado para realizar el TP
 - Las dificultades encontradas y cómo se resolvieron
 - También se debe detallar aquí el proceso de importación de los datos realizado
 - El informe debe tener como máximo 3 páginas

5. Evaluación

La evaluación del trabajo se llevará a cabo utilizando los parámetros establecidos en la rúbrica asociada a la actividad en el Campus.

Se tendrá en cuenta que las consultas, más allá del funcionamiento (lo cual es fundamental), sean genéricas.

Los docentes ejecutarán el proceso usando el conjunto de datos entregado, pero podrán también hacer pruebas con otros conjuntos de datos de similares características para evaluar el funcionamiento en distintos escenarios.

El informe deberá estar completo **y sin faltas de ortografía**.

En caso de que el trabajo no cumpliera los requisitos básicos para ser aprobado, los alumnos serán citados en la fecha de recuperatorio para defenderlo y corregir los errores detectados.