

BASES DE DATOS AVANZADAS

Enunciado de la Primera Práctica - Diseño de modelo relacional

Partimos de un conjunto de datos con temática de partidos de hockey sobre hielo y centrada en las estadísticas de los jugadores en cada uno de los partidos jugados (con datos sobre cada una de las jugadas realizadas) y en las estadísticas de cada uno de los equipos participantes en cada uno de esos partidos. Para poder manipular los datos de todas estadísticas, alguien ha creado (con mayor o menor acierto) dos tablas de datos: la primera centrada en las estadísticas de los jugadores y la segunda en las estadísticas de los equipos.

La tabla relacionada con las estadísticas de jugadores contiene una fila por cada jugada realizada por cada jugador en cada partido, con los datos de las asistencias, goles y tiros de cada jugador (assists, goals y shots) . Para ello también se almacena del jugador su identificador (player_id), nombre y apellido (firstName y lastName) y nacionalidad (nationality), el identificador de partido (game_id), temporada (season), fecha (date_time), equipo local (home_team_id), equipo visitante (away_team_id), goles del equipo local y goles del equipo visitante (home_goals y away_goals). También se almacenan el identificador de cada jugada, el identificador del equipo que la ha generado (team_id_for) y el que la ha sufrido (team_id_against), además de una descripción (event).

Esta relación se define como PrimeraTabla = < T1, L1> dónde:

T1 = {game_id (A), season (B), date_time (C), home_team_id (D), away_team_id (E), home_goals (F), away_goals (G), player_id (H), firstName (I), lastName (J), nationality (K), play_id (L), team_id_for (M), team_id_against (N), event (O), assists (P), goals (Q), shots (R) }

L1 = { $A \rightarrow C$, $A \rightarrow D$, $A \rightarrow E$, $A \rightarrow F$, $A \rightarrow G$, $C \rightarrow B$,

$H \rightarrow I$, $H \rightarrow J$, $H \rightarrow K$,

$L \rightarrow M$, $L \rightarrow N$, $L \rightarrow O$, $L \rightarrow A$,

$AH \rightarrow P$, $AH \rightarrow Q$, $AH \rightarrow R$ }

La tabla relacionada con las estadísticas de los equipos contiene una fila para los datos de cada partido de cada equipo. Para ello se almacena el identificador de partido (game_id), temporada (season), fecha (date_time), identificador del equipo local (home_team_id), identificador del equipo visitante (away_team_id), goles del equipo local y goles del equipo visitante (home_goals y away_goals). También se almacenan el identificador del equipo al que se refieren los datos del partido (team_id) y su nombre (teamName), además de los goles (tgoals), tiros (tshots) y golpes (thits) de dicho equipo en el partido indicado en los primeros atributos.

Esta relación se define como SegundaTabla = < T2, L2> dónde:

T2 = {game_id (A), season (B), date_time (C), home_team_id (D), away_team_id (E), home_goals (F), away_goals (G), team_id (S), teamName (T), tgoals (U), tshots (V), thits (W) }

L2 = { $A \rightarrow C$, $A \rightarrow D$, $A \rightarrow E$, $A \rightarrow F$, $A \rightarrow G$, $C \rightarrow B$,

$S \rightarrow T$, $AS \rightarrow U$, $AS \rightarrow V$, $AS \rightarrow W$ }

Se ejecutará el script “creacionBDyTablasycargaDatos.sql” que creará la base de datos “**practica1bda**” en MySQL, conteniendo estas dos tablas y sus correspondientes datos, que contienen las estadísticas de las jugadas llevadas a cabo durante algunos partidos celebrados en Enero y Mayo de 2013. **ATENCIÓN:** Los dos ficheros “csv” deben estar ubicados en el directorio “/EUI_PORTABLES/MySQL/Uploads/”

Se pide:

ATENCIÓN: A pesar de que se hayan creado dos tablas, por supuesto pertenecen a la misma base de datos y por tanto el objetivo final es tener una única base de datos con un conjunto de relaciones normalizadas y relacionadas entre ellas.

1. Determinar el grado de normalización de estas dos relaciones, **indicando las claves existentes**. Se asume que el conjunto de dependencias L1 y L2 representan, de manera independiente, un recubrimiento mínimo y no redundante.
2. Modificar con SQL el esquema de las dos tablas **EXCLUSIVAMENTE** para dar de alta la clave primaria elegida para cada una de ellas en el apartado anterior
3. Resolver en SQL las siguientes consultas:
 - a. ¿Cuántos partidos ha jugado en el año 2013 el equipo de nombre Blackhawks?
 - b. Por cada partido registrado en el mes de enero de 2013, mostrar el nombre de los equipos locales.
 - c. Por cada equipo, mostrar su identificador, nombre, y estadísticas (tgoals, tshots y thits) en los partidos jugados en 2013 como equipos locales.
 - d. Obtener, para cada partido que haya jugado el o los equipo(s) en los que haya jugado el jugador Adam McQuaid, el identificador del equipo, el nombre del equipo, el año, mes y día del partido y el identificador de cada jugada realizada junto con su descripción (event) para cada una de las jugadas que haya realizado el equipo en cada partido (event) ordenadas por la fecha del partido
4. Insertar mediante SQL, si es posible, (manteniendo las claves primarias creadas) los siguientes datos, explicando los problemas y anomalías encontradas en caso de haberlos:
 - a. Los datos del nuevo jugador que se llama Samuel Smithson y es canadiense (CAN) y al que se le ha asignado el identificador 5554441.
 - b. El partido de código 2012030999 entre los "Red Wings" (locales) y los "Penguins" que terminó 3 a 2 a favor de los locales.
 - c. Los datos de un nuevo equipo, denominado "Lasters" y cuyo código será 99.
5. Actualizar el resultado del partido de código '2012020053' ya que el partido terminó realmente 5 a 3 en vez de 6 a 4 como figura. Comentar problemas y anomalías si las hay.
6. A partir de febrero de 2013, el equipo denominado "Wild" pasa a denominarse "Wild Side". Reflejar este hecho en las tablas, explicando problemas y anomalías. Comentar si surge algún problema entre datos en las tablas y las dependencias funcionales.
7. Normalizar R hasta conseguir un conjunto de relaciones que estén en **FNBC** mediante el algoritmo de descomposición con LJ. Explicar todos los pasos dados.
8. Normalizar R mediante el algoritmo de síntesis para conseguir un conjunto de relaciones que estén, al menos, en 3FN. Explicar los pasos dados y comparar el resultado con el punto 7.
9. Dibujar un modelo entidad/relación que se corresponda con lo obtenido en el paso 7 y, si se encuentra alguna irregularidad, revisar el diseño obtenido realizando justificadamente los cambios que se estimen oportunos sin perder la normalización efectuada.
10. Escribir un script SQL para crear las tablas resultantes ya normalizadas y revisadas del punto 9, con sus correspondientes claves primarias y foráneas en la base de datos "practica1bda".
11. Insertar los datos en las tablas normalizadas desde las **tablas originales** (ignorar los cambios de los apartados 4, 5 y 6, volviendo a crear las dos tablas desde los ficheros originales si es necesario) en sus correspondientes tablas mediante las correspondientes sentencias "INSERT INTO ... SELECT ...".
12. Repetir los puntos 3 al 6 con la nueva base de datos. Comentar y justificar razonadamente las diferencias encontradas en cada punto.

Normas

- ✓ Se debe entregar una memoria explicativa que incluya la solución a los apartados solicitados, con una primera página que contenga los datos identificativos de los miembros del grupo (apellidos, nombre y número de matrícula)
- ✓ La práctica se realizará en grupos de dos alumnos. La fecha límite de entrega es la que consta en la tarea correspondiente en la plataforma de la asignatura.
- ✓ Se subirá a la plataforma **un único fichero que contenga todo lo solicitado** (en caso de subirse más de un fichero deberán ser comprimidos con zip). **La práctica se subirá a la plataforma exclusivamente en la cuenta de uno de los dos componentes del grupo y el nombre del fichero deberá incluir los dos apellidos de ambos alumnos, separados por un guion.**
- ✓ Esta práctica será objeto de defensa individualizada ya bien sea mediante examen escrito u oral.
- ✓ Para la realización de la práctica se dispone de todos los recursos necesarios en el centro de cálculo de la E.T.S.I.S.I. así como de toda la información necesaria en la plataforma moodle de la asignatura.