

BASES DE DATOS AVANZADAS

Enunciado de la Segunda Práctica – Técnicas de aceleración de consultas

Partimos de una base de datos sobre estadísticas de jugadas y partidos de hockey de la NHL. La base de datos es un **subconjunto** de la práctica 1 de la asignatura. Se van a considerar las siguientes tablas:

- **Game**, que contiene los datos de los partidos
- **Player**, que contiene datos de todos los jugadores.
- **Team**, que contiene los datos de los equipos
- **Player_plays**, que contiene los jugadores que han participado en cada una de las jugadas. **ATENCIÓN:** no hay más información acerca de las jugadas (se ha suprimido la tabla de **jugadas**).
- **Player_stats**, que contiene las estadísticas de cada jugador en cada partido. **ATENCIÓN:** Se considera que un jugador ha jugado un partido si existen estadísticas de ese jugador en el partido concreto. En caso contrario se asume que el jugador no ha jugado el partido.
- **Team_stats**, que contiene las estadísticas de cada equipo en cada partido. **ATENCIÓN:** Se considera que un equipo ha jugado un partido si existen estadísticas de ese equipo en el partido concreto. Evidentemente, ese equipo deberá aparecer como local o visitante en los datos del partido.

Se deberán realizar los siguientes pasos:

1. Se realizará la importación del esquema y datos de las cuatro tablas, mediante la opción “Data Import” del menú Server de MySQLWorkbench con el fichero “dumpBDAprac2.sql” disponible en el Moodle de la asignatura.
2. Estudio de planes de consulta e índices.
 - a. Crear una consulta SQL que permita obtener el número de partidos jugados por cada jugador para aquellos jugadores de nacionalidad estadounidense (USA) o canadiense (CAN) que hayan jugado más partidos que la media del número de partidos jugados por todos los jugadores. La consulta devolverá el **nombre** y **apellido** del jugador y su **edad actual**, así como el **número de partidos jugados**, pero el resultado estará **ordenado descendentemente** por **edad** e a igual edad por apellido seguido de nombre pero ascendentemente.
 - b. Estudiar el plan de consulta, tomando nota de los costes del mismo y comentarlo.
 - c. Crear las claves principales y foráneas mediante los ficheros script **CrearClavesPrimarias.sql** y **CrearClavesForaneas.sql**, y nuevamente estudiar el plan de consulta, comparando costes con el punto anterior.
 - d. Crear los índices que se estimen necesarios para mejorar la consulta.
 - e. Estudiar el plan de consulta con los nuevos índices y comparar resultados con los obtenidos en los puntos anteriores.
3. Optimización de consultas y estudio de planes de consulta
 - a. Eliminar los índices creados en el apartado anterior, manteniendo claves primarias y foráneas.
 - b. Definir en SQL al menos dos sentencias diferentes que permitan obtener los datos de los equipos que hayan jugado más partidos en los últimos seis meses del año 2017, devolviendo los atributos: identificador del equipo, nombre del equipo, número de partidos jugados, total de tiros (tshots) realizados en esos seis meses, media de goles del equipo por partido.
 - c. Crear los índices que permitan optimizar el coste de las consultas, analizando plan de consulta y coste para cada uno de los casos, justificando que solución es la más óptima.
4. Estudio de índices en actualizaciones.
 - a. Eliminar los índices creados en el apartado anterior manteniendo claves primarias y foráneas.
 - b. Crear un atributo en la tabla “player_stats” que se denomine “**puntuacion**” que sea un número entero. El contenido de dicho atributo, para cada estadística de cada partido será un valor que se

corresponderá con la suma de las cantidades de los atributos goals, shots y assists (es decir, en cada fila se sumarán los valores de esos atributos y se actualizará en el nuevo atributo “puntuación” con esa suma). Actualizar la tabla *player_stats* para que contenga dicha información tomando nota de su **tiempo de ejecución**.

- c. Volver a actualizar a *null* el atributo **puntuación** en todas las filas
- d. Por preverse un tipo de consultas con ordenación muy complicada, se desea crear un índice en la tabla “*player_stats*” por los atributos goals, shots, game_id, assists, player_id y puntuación, en ese orden. Crear el índice.
- e. Volver a ejecutar la sentencia que establece los valores del atributo puntuación a la suma comentada anteriormente. Comprobar tiempo de ejecución y compararlo razonadamente con el obtenido en el punto 4.b.

5. Desnormalización.

- a. Eliminar los índices creados en el apartado anterior, manteniendo claves primarias y foráneas.
- b. Crear una consulta que devuelva, para cada jugador su nombre, apellido, edad, total de asistencias (assists) y media de asistencias por partido, total de goles (goals) y media por partido para aquellos jugadores que tengan una edad entre 25 y 33 años.
- c. Aplicar la técnica de desnormalización que se considere más adecuada para acelerar la consulta del apartado 5.b, creando los scripts sql necesarios para modificar el esquema de la base de datos.
- d. Crear un script que actualice los datos implicados en la desnormalización.
- e. Crear los triggers necesarios para mantener actualizados los datos implicados en la desnormalización.
- f. Realizar la consulta 5.b sobre la base de datos desnormalizada. Estudiar coste y plan comparándolo con el obtenido en el apartado 5b.

6. Particionamiento.

- a. Eliminar las claves foráneas con el script proporcionado “**EliminarClavesForaneas.sql**”.
- b. Crear una consulta sql que obtenga, para cada jugador, su apellido, nombre, el número de partidos jugados y el número de jugadas realizadas por el jugador durante el año 2017. Estudiar coste y plan.
- c. Razonar justificadamente (**sin necesidad de implementarla realmente en SQL**) una variante de la estructura existente realizando un particionamiento horizontal de los datos con el objeto de mejorar el tipo de consultas (con diferentes años) que se ha realizado en el apartado 6.a
- d. Implementar en MySQL un particionamiento horizontal (mediante la sentencia **ALTER TABLE PARTITION**) que separe los datos de los partidos jugados en el año 2017 del resto. Realizar de nuevo la consulta 6.a y estudiar coste y plan comparándolo con lo obtenido en el apartado 6.a. Si se necesita modificar la clave primaria, hágase mediante la sentencia ALTER TABLE.

7. Establecer conclusiones generales de todo el presente trabajo.

Nota:

Para la medición de tiempos y costes de ejecución hay que tener en cuenta que tras la ejecución de una consulta SQL pueden quedar en cache los datos leídos, **para eliminar la cache** se puede utilizar la sentencia:

```
SET GLOBAL query_cache_type = 0;
```

O bien, poner *sql_no_cache* después de la palabra reservada SELECT.

Se puede, como **tercera opción**, deshabilitar la memoria caché en el sistema de **manera global** (lo que implica que va a afectar al rendimiento de todas las consultas sobre todas las bases de datos). Para ello en el fichero **my.ini** hay que poner un valor cero en la opción *query_cache_type*:

```
query_cache_type = 0;
```

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, **los mismos que para la primera práctica**. 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 fichero zip que contenga, por un lado la memoria de la práctica en formato pdf o doc y por otro todos los ficheros sql que solucionan los diferentes apartados listos para ser ejecutados. 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.