Práctica 2. Programación Web y Bases de Datos

Versión 1.0 15/10/2021

Introducción

Tras la realización de la Práctica 1, el espíritu de esta práctica es la introducción de una base de datos para que aquellos elementos persistentes de forma dinámica en el sistema, como la información personal de los clientes, su saldo para compras y programas de fidelización y su historial de compras, así como toda la información relativa a las películas, actores, directores, etc. (catálogo), en vez de estar repartidos en multitud de directorios y archivos, se guarden en un único lugar, y sean accesibles y manipulables con todas las prestaciones y posibilidades que provee un sistema general de base de datos (SGBD, DBMS en inglés).

Objetivos docentes

Tras la realización de la Práctica 2, el estudiante debe ser capaz de:

- Entender los conceptos involucrados en el diseño de una base de datos relacional, en especial las relaciones que se deben establecer entre las distintas tablas.
- Realizar consultas de diferente complejidad sobre bases de datos relacionales.
- Entender y manejar con agilidad los triggers en una base datos.
- Entender y ser capaz de implementar Funciones y Procedimientos Almacenados.
- Acceder a bases de datos desde el lenguaje Python con la interfaz SQLAlchemy, realizando actualizaciones y consultas bajo demanda del usuario.

Base de Datos

Para facilitar la tarea, se proporciona una base de datos volcada en un fichero de texto (fichero incluido en **dump_v1.4.sql.gz**) que servirá de punto de partida para implementar lo solicitado en la práctica. Para cargar este volcado, suponiendo que el usuario alumnodb ya existe en PostgreSQL (consultar las instrucciones de la plataforma), podemos ejecutar los siguientes comandos:

```
createdb -U alumnodb si1
gunzip -c dump_v1.4.sql.gz | psql -U alumnodb si1
```

En los anteriores comandos, **si1** es el nombre que se le pone a la base de datos. **Debe utilizarse siempre este nombre para facilitar la corrección del código entregado**. El primer comando crea la propia base de datos, mientras que el segundo descomprime el volcado y rellena la base de datos a partir de lo especificado en él.

Grosso modo, la base de datos está organizada en las siguientes tablas:

- *imdb_movies*: representa las distintas películas que se pueden comprar.
- *products*: almacena los distintos productos que pueden existir de la misma película con sus precios.
- *inventory*: indica las ventas y el stock de cada producto.
- *customers*: guarda la información relativa a los diferentes clientes.

- *orderdetail*: detalla las compras concretas que han ido realizando los clientes a través de la web.
- *orders*: agrupa los productos de *orderdetail* en pedidos (carritos finalizados, es decir, que se han pagado) propios de cada usuario.

Algunas características adicionales de esta base de datos son:

- Un pedido en curso (cesta o carrito), se caracteriza por tener un valor NULL (valor reservado de SQL, no una cadena de caracteres) en la columna *status* de la tabla *orders*. Por ello, sólo puede haber como máximo un registro de la tabla *orders* con *status* NULL para un cliente dado.
- El pedido pasa sucesivamente por los siguientes valores de *status*: NULL, '*Paid*', '*Processed*', '*Shipped*'.
- La columna *sales* de la tabla *inventory* contiene el número acumulado de artículos vendidos de un producto.

Tareas de la práctica

Se solicitan a continuación una serie de tareas e informes sobre la base de datos. Si se encuentra que los datos existentes no son suficientes para dar respuesta a estas peticiones, el alumno deberá incorporar a la misma la información extra necesaria.

Diseño de la BD

a) En primer lugar, se deberá efectuar un proceso de ingeniería inversa sobre la base de datos suministrada. En particular, se deberá:

- Obtener el diagrama entidad-relación correspondiente. Se puede realizar manualmente o utilizando las herramientas que sean necesarias.
- Discutir el diseño de la base de datos, ventajas y desventajas de las decisiones implícitas en el mismo.
- Completar aquellos aspectos que se consideren necesarios, tales como restricciones, claves extranjeras, cambios en cascada, etc. Todos estos cambios se incorporarán en un único script, **actualiza.sql**.
- Añadir a **actualiza.sql**:
  - Una nueva tabla 'alerts' para guardar los momentos (fecha/hora) en los que el stock de un producto (inventario) llega a cero.
  - Un campo numérico 'loyalty' en la tabla 'customers', con valor por defecto cero, para guardar los puntos de fidelización de los clientes.
  - Un campo 'balance' en la tabla 'customers', para guardar el saldo de los clientes.
- Crear un procedimiento que inicialice el campo 'balance' de la tabla 'customers' a un número aleatorio entre 0 y N, con signatura:

```
function setCustomersBalance(IN initialBalance bigint);
```

- Añadir a **actualiza.sql** una llamada a dicho procedimiento, con N = 100.

Se deberá entregar, al menos, el **diagrama E-R final**, tras los cambios del script actualiza.sql.
A la hora de elaborar este diagrama conviene revisar:

- claves primarias

- claves externas o extranjeras
- qué tablas son entidades, cuáles relaciones y cuáles atributos
- cardinalidad
- entidades débiles
- atributos multivaluados
- atributos derivados
- participación total

b) Sabiendo que los precios de las películas se han ido incrementando un 2% anualmente, elaborar la consulta **setPrice.sql** que complete la columna 'price' de la tabla 'orderdetail', sabiendo que el precio actual es el de la tabla 'products'.

c) Una vez se disponga de esta información, realizar un **procedimiento almacenado**, **setOrderAmount**, que complete las columnas 'netamount' (suma de los precios de las películas del pedido) y 'totalamount' ('netamount' más impuestos) de la tabla 'orders' cuando éstas no contengan ningún valor. Invocadlo en **actualiza.sql** para realizar una carga inicial.

El procedimiento almacenado no lleva parámetros.

d) Realizar una función postgreSQL, **getTopSales**, que reciba como argumentos dos años diferentes y devuelva las películas que más se han vendido entre esos dos años, una por año, ordenadas de mayor a menor por número de ventas.

Year	Film	Sales
1897	El Gran Dictador	12
1912	Superman	4
...		

(Sólo es un ejemplo, no son resultados reales)

La signatura de la función es:

```
function getTopSales(year1 INT, year2 INT,
                     OUT Year INT, OUT Film CHAR, OUT sales bigint);
```

e) Realizar una función postgreSQL, **getTopActors**, que reciba un parámetro "género" y que devuelva los actores o actrices que más veces han actuado en dicho género, ordenados de más actuaciones a menos, **siempre que hayan trabajado en más de 4 películas de ese género**, con información de la película en la que debutaron para ese género, el año de esa película y quién (o quiénes) dirigió esa película (pueden aparecer varios registros para el mismo actor porque hizo varias películas su primer año, varios directores para la misma película, etc.)

Actor	Num	Debut	Film	Director
Costner, Kevin	6	1985	American Flyers (1985)	Badham, John
Gibbs, Keith	5	1994	Air Up There, The (1994)	Glaser, Paul Michael

Gibbs, Keith	5	1994	Blue Chips (1994)	Friedkin, William
...				

(Sólo es un ejemplo, no son resultados reales)

La signatura de la función es:

```
function getTopActors(genre CHAR, OUT Actor char, OUT Num INT, OUT Debut INT,
                      OUT Film CHAR, OUT Director CHAR);
```

Integridad de los datos

f) Para garantizar la integridad de los datos (los valores posibles de las columnas), crear las tablas correspondientes y convertir los atributos multivaluados 'moviecountries', 'moviegenres' y 'movielanguages' en relaciones entre la tabla 'movies' y las tablas creadas. Estos cambios también se incorporarán al script **actualiza.sql**.

g) Realizar un trigger, **updOrders**, que actualice la información de la tabla 'orders' cuando se añada, actualice o elimine un artículo del carrito.

h) Realizar un trigger, **updInventoryAndCustomer**, explicando de forma adecuada en la memoria **cuándo dispararlo**, que realice las siguientes tareas:

- actualice las tablas 'orders' e 'inventory'
- cree una alerta en la tabla 'alertas' para aquellos productos comprados cuya cantidad en stock en el inventario llega a cero
- sume los puntos de fidelización conseguidos por la compra en la tabla 'customers'
- descuente en la tabla 'customers' el precio total de la compra

Integración en el portal

i) Incorporar la tabla resultante en el apartado e) anterior, para el género "Action" y limitado a 10 filas, a la página inicial de bienvenida. **Se puntuará mejor** si se puede seleccionar el género y la cantidad de películas de una lista/cuadro y si el nombre de la película es un enlace que lleva a información de la misma.

j) Implementar en el sitio web el registro y la validación de usuario (login) usando la tabla 'customers'. No es preciso cifrar las contraseñas.

k) Implementar el resto de páginas con contenidos de películas. En particular, **para la pantalla inicial de bienvenida**, tened en cuenta que las películas que se muestran y compran **son las de la base de datos** (no las del catálogo de la práctica anterior). Por tanto, en ausencia de filtros y buscadores (o incluso con ellos), es conveniente hacer una selección inicial **limitada** de películas para mostrar en dicha pantalla inicial (fija, por fecha, las que tienen imágenes, aleatoria, con o sin opción LIMIT, etc.) pues si no la carga puede ser lenta, haciendo el sitio poco "user-friendly".

l) Implementar en el sitio web la funcionalidad de carrito, usando las tablas de la base de datos apropiadas. Se puede mantener la funcionalidad de carrito por sesión de la práctica 1 para el caso de navegación por el sitio web sin haber hecho login.

(***Nota***: **Con login hecho** se debe usar la BD para almacenar el carrito desde el primer artículo que se introduzca en él; no es válido ir guardando el carrito en la sesión y volcarlo a la BD al hacer *logout*)

El código Python de acceso a la base de datos deberá usar la interfaz implementada en SQLAlchemy (https://www.sqlalchemy.org/).

Entregables

Como **resultado** de la práctica se entregará:

- Código fuente:
    - Parte SQL: como se ha especificado más arriba, los ficheros SQL que hay que entregar son (los nombres de los archivos se corresponden con los de los triggers, funciones, etc. que contienen):
        - **actualiza.sql:** script que transforma la base de datos suministrada en la de trabajo de la práctica, tras corregir errores, incorporar mejoras, etc.
        - **setPrice.sql**
        - **setOrderAmount.sql**
        - **getTopSales.sql**
        - **getTopActors.sql**
        - **updOrders.sql**
        - **updInventoryAndCustomer.sql**
    - **La base de datos como tal (dump) no debe entregarse, debido a su tamaño.** Se entiende que el script 'actualiza.sql' es el que transforma el dump suministrado originalmente en la base de datos propia de cada pareja.
    - Parte web: Python y auxiliares (CSS, JS, HTML…). Recordar **no entregar el entorno de python si1pyenv**
- Memoria en la que se incluirá:
    - El diagrama entidad-relación de la base de datos resultante tras aplicar el script 'actualiza.sql' y los cambios introducidos, con su justificación.
    - El análisis de la solución dada a las consultas, triggers y funciones solicitados.
    - Las evidencias de los resultados obtenidos.
    - Cualquier mejora o propuesta de implantación que el alumno considere.

IMPORTANTE: Si es necesario incorporar datos adicionales para completar las tablas del modelo y para hacer pruebas, estos cambios deben ser documentados debidamente en la memoria entregada.

Entrega

La fecha concreta y normas de entrega de las prácticas se encuentran en Moodle.

Evaluación

Se valorará muy positivamente cualquier mejora o añadido adicional que aporte el alumno, discusión sobre la eficiencia del método empleado, etc.

Revisad que el formato de las sentencias SQL de consulta se adecúe a las Normas de Prácticas.

Consejos

SQL

Si se hace '*update mi_tabla set mi_columna = (select …* ' revisar si se puede evitar vía '*update …. from …*'. De igual forma, si se hace un '*delete*' similar (que use un '*select*') revisar '*delete …. using …*'.

Si se hace ' *… from (select …*' revisar si de verdad es necesario ese select o se puede sustituir con un join.

En las funciones de base de datos, intentar evitar los bucles y usar las capacidades de las sentencias SQL: por ejemplo, una sola sentencia '*update*' puede cambiar muchos registros, no hace falta un bucle que los recorra todos.

Si la función de un trigger no hace referencia a '*NEW*' ni a '*OLD*', casi seguro que está mal, es muy ineficiente (en esta práctica, seguro que es así).

Al programar la función de un trigger hay que tener siempre cuidado de no ejecutar código cuando no se debe, por ejemplo, porque se ha actualizado sólo una columna que no es la relevante de cara a la ejecución del trigger.

Recordad que, en un trigger, la variable TG_OP indica si el trigger se disparó por un INSERT, UPDATE o DELETE.

Recordad que un diagrama E-R no es un diagrama de diseño de una base de datos, con sus tablas y relaciones.

El uso de *pgadmin3* puede dar problemas al guardar los ficheros. En particular, añade una cabecera de la UTF8 (‘BOM’) que afecta a la posterior ejecución de los scripts de BD. Se puede utilizar un editor como *geany* para eliminarlos (opción Document->uncheck ‘Write Unicode BOM’.

SQLAlchemy

Se suministra un ejemplo de recuperación de datos de la BD en example-web-alchemy-v<x.y>.zip. Se recomienda revisarlo detenidamente.

SQLAlchemy provee dos interfaces de acceso a la BD: vía objetos (ORM) o vía funciones (Core). Se recomienda usar SQLAlchemy Core.

En la actualidad SQLAlchemy se encuentra en período de transición a la versión 2.0. Se recomienda encarecidamente utilizar la versión 1.x de dicho producto. La última disponible es la 1.4, mientras que en los laboratorios de la EPS está instalada la 1.1. Puede haber pequeñas diferencias entre versiones, así que es importante tener en cuenta la documentación.

Para ejecutar sentencias SQL se puede usar el método *execute* de una *connection* (ver https://docs.sqlalchemy.org/en/14/core/connections.html#sqlalchemy.engine.Connection).

El resultado de una consulta se devuelve vía el objeto *Result* (https://docs.sqlalchemy.org/en/14/core/connections.html#sqlalchemy.engine.Result), que contiene las filas resultantes.

Cada fila del resultado es un objeto *Row* (https://docs.sqlalchemy.org/en/14/core/connections.html#sqlalchemy.engine.Row). Los valores de cada columna se obtienen vía *items()*, que devuelve un array de tuplas clave-valor.

También se puede recuperar la información a través de los métodos *first()*, *fetchone()*, *fetchall()*, … del objeto *Result*, que funcionan de forma análoga a los cursores.

Recordamos que, para utilizar las bibliotecas que más nos convengan, **siempre podemos instalarlas en nuestro entorno virtual de Python**. En ese caso debemos indicarlo en la memoria e incluir un fichero requirements.txt en nuestro proyecto.

Por ejemplo, si queremos instalar la versión 1.4 de SQLAlchemy:

```
$ virtualenv -p python3 si1pyenv
...
$ source si1pyenv/bin/activate
$ pip3 install SQLAlchemy==1.4
...
```

Bibliografía

[1] Documentación de postgreSQL: https://www.postgresql.org/docs/10/index.html
[2] Acceso a bases de datos en Python usando SQLAlchemy: https://www.sqlalchemy.org/
[3] PgAdmin: http://www.pgadmin.org/