

Práctica 3: Diseño de Bases de Datos Federadas

Bases de Datos 2

Grado en Ingeniería Informática
Departamento de Informática e Ingeniería de Sistemas
Universidad de Zaragoza

1.- Objetivos

En esta práctica el alumnado se enfrentará a un problema de integración de bases de datos para definir una base de datos federada. En mayor detalle, los objetivos de la práctica son:

1. Conectarse a una base de datos (BD) Oracle proporcionada y analizarla (a partir de la información disponible en el diccionario de datos y de los datos disponibles) para tratar de inferir su esquema conceptual.
2. Integrar el esquema de dicha BD con el desarrollado previamente en la práctica 1, ofreciendo un esquema global.
3. Integrar dos BD en PostgreSQL.

Para ello es interesante considerar, al menos, la utilización de *DATABASE LINK* de Oracle y de *dblink* de PostgreSQL. Asimismo, es interesante valorar otras opciones que pudieran existir.

2.- Contexto del Problema

En la práctica 1+2 de la asignatura hemos abordado el diseño de una base de datos relacional para un banco, que en adelante consideraremos que se llamaba *Banquito*.

Parte 1: objetivos 1 y 2

Ahora *Banquito* ha sido integrado en *Banquete*, un banco más grande que también contaba con su propia base de datos para almacenar su información. Nuestro objetivo es diseñar e implementar un esquema global que permita ver ambas bases de datos relacionales de manera uniforme, ya que de momento se prevé que ambas bases de datos mantendrán su autonomía. Afortunadamente, ambas bases de datos estaban implementadas sobre Oracle y ambas siguen el modelo relacional, lo que se supone que nos va a facilitar considerablemente nuestra tarea. No obstante, deben tenerse en cuenta los siguientes aspectos:

- No se dispone de documentación acerca del esquema de la base de datos a integrar. Por tanto, tenemos que recurrir a consultar el diccionario de datos para tratar de inferir el esquema.
- El esquema global debe recoger toda la riqueza semántica de los esquemas subyacentes de la mejor manera posible, de modo que los usuarios de las bases de datos locales (y en particular los de *Banquete*) se sientan lo más cómodos posible cuando tengan que consultar el esquema global.

- No es posible modificar el diseño de los esquemas locales. No obstante, en el proceso de integración podemos (y debemos) detectar aspectos mejorables en dichos esquemas; en ese caso, los anotaremos y los comentaremos en la memoria de la práctica, para que puedan tenerse en cuenta en futuras mejoras del sistema de información del banco integrado (se comunicarían dichas sugerencias y mejoras a los administradores/diseñadores de las fuentes de datos locales, para su consideración).

Por el momento, se nos pide que realicemos la integración con una copia de la base de datos de producción de *Banquete*, que contiene únicamente un subconjunto de datos de prueba. Posteriormente, cuando todo esté en marcha, y tras un periodo de prueba, se realizará la carga del conjunto de datos completo; esta labor queda fuera del alcance de la práctica. No obstante, para realizar la integración debemos asumir que los datos almacenados son válidos a todos los efectos.

Parte 2: objetivo 3

En esta segunda parte de la práctica, se pretende hacer algo similar, pero esta vez integrando dos esquemas de bases de datos implementadas en PostgreSQL, en lugar de implementadas en Oracle.

Para esta segunda parte de la práctica, debéis: 1) plantear un problema original de diseño de bases de datos (escribir un enunciado de un problema), 2) realizar el diseño conceptual de dicha base de datos (esquema E/R), 3) realizar el diseño lógico de dicha base de datos e implementarlo en PostgreSQL, 4) realizar un segundo diseño lógico alternativo e implementarlo también en PostgreSQL (debe realizarse un diseño lógico diferente al anterior: forzando heterogeneidad de nombrado, diferentes formas de traducir especializaciones/generalizaciones del esquema E/R al esquema relacional, etc.), y 5) integrar ambos esquemas. Se debe definir el segundo esquema lógico de modo que la integración no resulte trivial sino que plantee diversas dificultades; para ello, es posible considerar un segundo esquema que modele situaciones ligeramente diferentes a las del primer esquema (es decir, considerar dos enunciados de problema relacionados en lugar de un único enunciado).

3.- Aclaraciones

Es conveniente realizar las siguientes aclaraciones:

- Es importante extraer conclusiones: resultados obtenidos, indagaciones realizadas, dificultades encontradas, valoración de los aprendizajes o mejoras de habilidades conseguidas, opinión personal, etc.
 - En particular, para esta práctica es relevante:
 - Explicar claramente la forma en la que habéis deducido el esquema conceptual de la base de datos que no habéis diseñado vosotros (esquema externo a integrar).

- Explicar y justificar la forma en la que se construye el esquema global, así como enfatizar las distintas dificultades de integración que puedan aparecer.
 - Analizar y comentar la posibilidad y posible conveniencia de permitir, además de consultas, actualizaciones sobre el esquema global y sus implicaciones.
 - Recoger sugerencias de mejora y posibles problemas o deficiencias de diseño del esquema de la base de datos de *Banquete*.
- Algunas notas concretas de utilidad para esta práctica:
 - Hay un DATABASE LINK creado para la práctica 3 en cada uno de los servidores de bases de datos de Oracle actualmente accesibles desde el laboratorio, que apuntan a la base de datos a integrar (en realidad cada uno apunta a una base de datos diferente, alojada en un servidor distinto, pero los contenidos de ambas bases de datos son equivalentes). Para acceder al enlace alojado en el servidor Oracle 9i (hendrix) podéis referenciar el enlace mediante **SCHEMA2BD2** o **Schema2BD2**; para acceder al enlace alojado en el servidor Oracle 18 (danae04), podéis referenciar el enlace mediante **SCHEMA2BD2**.
 - Para descubrir el esquema de la base de datos a integrar, se recomienda consulta el apartado “Metadatos en Oracle” en las notas de Oracle proporcionadas en Moodle.
 - La posibilidad de manipular (insertar o modificar) los datos de la base de datos a integrar en el caso de Oracle podría “despistar” a otros/as compañeros/as, que tienen que integrar la misma base de datos. Por tanto, incluso si es posible, no debe intentarse esta posibilidad directamente sobre la BD de Oracle proporcionada por el profesor. Dicho de otro modo, no se debe intentar modificar datos en la fuente de datos local de *Banquete*. Sin embargo, sí que es interesante probarlo con una réplica de dicha BD o con otro ejemplo.
 - Como parte de la práctica es interesante que detectéis cosas que mejoraríais / corregiríais en el esquema a integrar, aunque como se indica en el guión no podáis realizar modificaciones en el mismo.
 - Un código de cuenta corriente (CCC) consta de 20 dígitos (http://es.wikipedia.org/wiki/C%C3%B3digo_cuenta_cliente). Si os encontráis con sólo 18 en la BD a integrar podéis suponer que los 2 que faltan son los correspondientes a los dígitos de control.
 - En caso de problemas con el enlace Schema2BD2 mencionado, avisad al profesor de inmediato y continuad entre tanto con otras partes de la práctica y pruebas con las máquinas virtuales.

4.- Entrega de la Práctica

Se debe entregar un fichero zip denominado p3-<nip>.zip (donde <nip> representa el NIP del coordinador/responsable del grupo) con el siguiente contenido:

1. Fichero autores.txt: contendrá el nombre y apellido de los autores de la práctica y sus NIP.
2. Directorio *fuentes*: contendrá todo el código desarrollado para la práctica:

- Fichero bd-bancaria-oracle-p1.sql: código SQL para la creación del esquema de la BD relacional de Oracle definido en la práctica anterior (*parte 1 de la práctica 3*).
- Fichero bd-bancaria-oracle-integrada.sql: código SQL para la creación del esquema global de la BD relacional de Oracle (*parte 1 de la práctica 3*).
- Fichero bd-bancaria-oracle-integrada-pruebas.sql: código SQL con las pruebas realizadas sobre el esquema global definido para la BD de Oracle: consultas de ejemplo, etc. (*parte 1 de la práctica 3*).
- Fichero bd-bancaria-postgresql-esquema1-p3.sql: código SQL para el primer esquema de la BD de PostgreSQL (*parte 2 de la práctica 3*).
- Fichero bd-bancaria-postgresql-esquema2-p3.sql: código SQL para el segundo esquema de la BD de PostgreSQL (*parte 2 de la práctica 3*).
- Fichero bd-bancaria-postgresql-integrada.sql: código SQL para la creación del esquema global de la BD de PostgreSQL (*parte 2 de la práctica 3*).
- Fichero bd-bancaria-postgresql-integrada-pruebas.sql: código SQL con las pruebas realizadas sobre el esquema global definido para la BD de PostgreSQL: consultas de ejemplo, etc. (*parte 2 de la práctica 3*).

Importante: cada uno de los ficheros .sql indicados deberá poder ejecutarse sin problemas como fichero de script con independencia del estado inicial de la base de datos destino, de forma que en el script hay que asegurarse de limpiar adecuadamente el espacio de tablas de la base de datos para evitar conflictos con objetos del mismo nombre que pudieran existir previamente (tablas, restricciones, etc.): como ejemplo de prueba, debe ser posible ejecutar dos veces seguidas el script y no producirse errores.

3. Fichero memoria-p3.pdf, conteniendo la memoria de la práctica, de acuerdo con las instrucciones generales de elaboración de prácticas. La memoria de esta práctica deberá contener al menos las siguientes secciones:
 - “Esfuerzos invertidos”: horas invertidas por cada integrante del grupo en las distintas tareas.
 - “Parte 1 – Esquema conceptual y lógico de la base de datos relacional diseñada en la práctica anterior”.
 - “Parte 1 – Determinación del esquema lógico y conceptual de la base de datos Oracle a integrar”, incluyendo las operaciones realizadas para consultar el catálogo.
 - “Parte 1 – Mejoras sugeridas para la base de datos a integrar”: descripción de posibles problemas o aspectos mejorables en el diseño y/o datos existentes en la base de datos Oracle a integrar.
 - “Parte 1 – Definición e implementación del esquema global en Oracle”: esquema E/R global y su implementación.
 - “Parte 2 – Enunciado de un problema de diseño de bases de datos”.
 - “Parte 2 – Esquema conceptual para el problema enunciado” (esquema E/R).
 - “Parte 2 – Esquema lógico 1 para el problema enunciado”.
 - “Parte 2 – Esquema lógico 2 para el problema enunciado”.

- “Parte 2 – Esquema global y su implementación en PostgreSQL”, incluyendo pruebas que muestren que la integración de las dos bases de datos implementadas en PostgreSQL es correcta.
- “Actualizaciones de datos sobre el esquema global”: comentar en qué medida es o no posible o interesante insertar o modificar datos accediendo directamente al esquema global (en Oracle y PostgreSQL), así como las posibles implicaciones, incluyendo alguna prueba para mostrarlo.

Instrucciones de envío:

El fichero anterior se someterá a través de *Moodle* (se habilitará una opción a tal efecto) utilizando la cuenta del coordinador/responsable del grupo.

Hay que asegurarse de que la práctica funciona correctamente y que será fácilmente reproducible (incluir ficheros de scripts para Windows y Linux, si resulta aplicable). También es importante someter código limpio (donde se ha evitado introducir mensajes de depuración y comentarios que no proporcionan información relevante).

Fecha límite de entrega:

La fecha límite de entrega será el día anterior a la siguiente sesión de prácticas.