

Práctica 1: Instalación y Configuración de Sistemas Gestores de Bases de Datos

Bases de Datos 2

Grado en Ingeniería Informática
Departamento de Informática e Ingeniería de Sistemas
Escuela de Ingeniería y Arquitectura
Universidad de Zaragoza

1.- Objetivos

En esta práctica vamos a proceder a la instalación y configuración de distintos Sistemas Gestores de Bases de Datos (SGBD), sobre los que posteriormente iremos desarrollando distintas tareas en las sucesivas prácticas de la asignatura. Como parte de esta práctica, trabajaremos también conceptos básicos de administración y configuración de seguridad de los distintos SGBD.

En mayor detalle, los objetivos de la práctica son:

1. Instalar diversos Sistemas Gestores de Bases de Datos. Los distintos SGBD que utilizaremos soportan diferentes sistemas operativos y presentan diferencias en su modo de instalación. Además, pueden plantearse diversos procedimientos para su instalación. Por ejemplo, podemos instalar SGBD sobre nuestro sistema operativo anfitrión (Linux, Windows, MacOS) como instalaríamos cualquier otra aplicación; podemos instalarlos sobre máquinas virtuales Linux, con la ventaja de homogeneizar los mecanismos de instalación, e independizarlos de nuestro sistema anfitrión; o podemos, finalmente, utilizar contenedores Docker para esta tarea de instalación, lo cual simplifica considerablemente la tarea de desplegar y arrancar un SGBD, siempre que nuestro sistema operativo soporte adecuadamente este entorno de contenedores.
2. Realizar la adecuada configuración de seguridad de los distintos SGBD instalados, creando *superusuarios* y diferentes roles / usuarios, con diferentes permisos (creación de objetos tales como tablas, gestión de datos, acceso de sólo lectura, etc.).
3. Realizar la configuración del espacio de datos de los distintos SGBD, creando esquemas / bases de datos (según el gestor utilizado) y tablas de ejemplo, verificando las condiciones de seguridad establecidas (de acuerdo a lo comentado en el objetivo 2).
4. Generar, insertar y consultar un mínimo conjunto de datos mediante instrucciones SQL (o similares), desde *CLI (Command-Line Interface)* o desde entornos de desarrollo de bases de datos (*IDEs* de acceso a base de datos), para comprobar el correcto funcionamiento de los gestores instalados.
5. Desarrollar la capacidad de buscar documentación relevante sobre los gestores, experimentar con los mismos y resolver los problemas que puedan surgir de forma autónoma.

2.- Instalación de Sistemas Gestores de Bases de Datos

En la actualidad, existe una gran cantidad de sistemas gestores de bases de datos, tanto relacionales como NoSQL (entre estos últimos, sistemas basados en clave/valor, en columnas, grafos, documentales o mixtos, en una clasificación sencilla de los mismos). Incluso, entre los SGBD relacionales, aunque todos ellos implementan el mismo modelo lógico y son compatibles en gran medida a través del lenguaje de definición y consulta SQL, no son todos idénticos, y la nomenclatura utilizada para denominar los distintos elementos, o la forma de configurar el espacio de datos, pueden ser ligeramente distintos.

En esta práctica vamos a instalar, configurar y explorar los siguientes SGBD:

- PostgreSQL.
- Oracle XE.
- Apache Cassandra.
- Apache HBase (ecosistema Hadoop).
- IBM DB2.

La mayoría de los SGBD anteriores serán utilizados en las siguientes prácticas, por lo que es importante instalarlos y configurarlos adecuadamente.

Estos gestores deberán instalarse utilizando alguna de las siguientes opciones:

- Instalación nativa en el propio sistema operativo de la máquina anfitriona. Es una opción posible, aunque en general poco recomendable, salvo que nos interese tener un gestor de base de datos instalado localmente y configurado para poder trabajar con él de forma habitual. En el caso de las prácticas esta opción sería únicamente posible si se trabajara con un ordenador personal, ya que no se tienen permisos para realizar instalaciones en máquinas del laboratorio.
- Instalación en máquina virtual. Podemos crear una máquina virtual Linux sobre VirtualBox (o cualquier otro hipervisor, como VMware o Xen), e instalar estos SGBD en dicha máquina virtual, o incluso configurar una máquina virtual separada para cada uno de los SGBD, y así evitar conflictos e interferencias.
- Instalación mediante contenedores Docker. Necesitaremos tener instalado el servicio Docker en un ordenador personal o bien en una máquina virtual Linux (si nuestro equipo tiene instalado Windows como sistema base, siempre es preferible utilizar una máquina virtual Linux para instalar Docker, aunque Windows 11 ha mejorado considerablemente el soporte de Docker y procesos Linux). Todos los SGBD con los que vamos a trabajar tienen al menos una imagen oficial en Docker. Tendremos que revisar cuidadosamente, eso sí, la forma de parametrizar la creación de nuestros contenedores, para que cumplan los requisitos de funcionalidad y seguridad requeridos.

3.- Configuración de los SGBD

Con los distintos SGBD instalados, debemos realizar al menos la siguiente configuración:

- Creación de un *superusuario*, con credenciales seguras, y verificación de que podemos conectarnos con dicho usuario
- Creación de la estructura básica del espacio de datos: base de datos/esquema/*keyspace*/tabla/*column_family*/... (la nomenclatura y estructura depende de cada SGBD).
- Creación de usuarios y roles con distinto acceso sobre los elementos del espacio de datos. En cada SGBD será necesario, al menos, tener un usuario (no *superusuario*) con capacidad para escribir (insertar/modificar) datos sobre las tablas creadas, y otro usuario con permiso exclusivamente para leer datos. En aquellos SGBD que distingan el objeto *User* del objeto *Role*, la asignación de permisos se hará a nivel de *Role*.
- El SGBD instalado deberá permitir el acceso remoto a los datos desde otro ordenador conectado a la misma red local. Para ello, se deberá comprobar que es posible conectarse al SGBD desde un cliente instalado en otro ordenador conectado en la misma red local que el equipo donde se ha instalado el SGBD y que no es posible el acceso desde fuera de dicha red. Si algún SGBD no permitiera una configuración tan fina, se deberá comentar y analizar en la memoria entregada.

4.- Poblado y Prueba de Acceso

Una vez instalado y configurado adecuadamente cada SGBD, se deberán ejecutar los comandos SQL (o lenguaje equivalente) necesarios para crear y consultar un número mínimo de registros en las tablas de ejemplo creadas.

3.- Aclaraciones

Es conveniente realizar las siguientes aclaraciones:

- Se proporcionarán diversas instrucciones en las clases de laboratorio y a través de *Moodle* (por ejemplo, acerca del uso de *PostgreSQL*, *Oracle* o *Cassandra*, así como instrucciones de ayuda para la instalación de algunos gestores o de manejo de máquinas virtuales).
- Es importante extraer conclusiones: resultados obtenidos, indagaciones realizadas, diferencias observadas entre los distintos Sistemas Gestores de Bases de Datos (SGBD), dificultades encontradas, valoración de los aprendizajes o mejoras de habilidades conseguidas, opinión personal, etc. En la medida de lo posible, y dentro de lo razonable, la memoria de prácticas contendrá el suficiente nivel de detalle como para poder reproducir el trabajo realizado.

4.- Entrega de la Práctica

Se debe entregar un fichero zip denominado p1-<nip>.zip (donde <nip> representa el NIP del coordinador/responsable del grupo) con el siguiente contenido:

- Fichero autores.txt: contendrá el nombre y apellido de los autores de la práctica y sus NIP.
- Directorio *fuentes*: contendrá todo el código desarrollado para la práctica:
 - Ficheros con las instrucciones o scripts ejecutados para la instalación de los distintos SGBDs.
 - Ficheros con las sentencias (DDL y DML) utilizadas para la configuración de los entornos de seguridad, modelos de datos y registros creados o consultados.

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.

- Fichero memoria-p1.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.
 - **“Configuración del entorno de trabajo (máquina virtual / uso de Docker)”** (sección a incluir si se ha utilizado alguna de estas 2 opciones de trabajo): descripción de los pasos seguidos para configurar la máquina virtual (instalación de utilidades, configuración del teclado, otros ajustes, etc.). No debe replicarse información de configuración que ya proporcione directamente el profesorado a través de Moodle, salvo que sea para complementar o precisar dicha información.
 - **“Instalación y Administración Básica de los SGBD”**: descripción detallada del procedimiento de instalación de los distintos SGBD propuestos (concretamente, Oracle, PostgreSQL, Cassandra, HBase y DB2), así como de las pruebas realizadas para asegurar que las instalaciones se realizaron de forma correcta. Se deberán indicar también los pasos básicos de administración seguidos para la puesta en marcha: crear una base de datos, arrancar y parar al gestor, crear los usuarios y otorgarles permisos para que únicamente ellos puedan utilizar la base de datos, configurar las formas de acceso permitidas, etc.
 - **“Comentarios Acerca de las Licencias”**: aspectos más relevantes de las licencias de uso de los distintos SGBD, citando adecuadamente las porciones más relevantes de texto de las licencias e indicando su implicación.

- **“Generación de Datos y Pruebas”**: código SQL para insertar datos de prueba y consultas de ejemplo para probar las bases de datos construidas. En este apartado se deben analizar estrategias y herramientas para la generación de datos de ejemplo.

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). 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 al de realización de la práctica 2.