

Bases de datos

En Humanidades digitales, como en el resto de áreas de conocimiento, es necesario organizar la información de modo que su acceso sea más fácil. No hace falta dar muchas razones de ello, basta con pensar en las tareas más habituales en las que tenemos que procesar datos para que la evidencia se imponga por sí sola. Las cuestiones que trataremos aquí son esencialmente dos: los tipos de bases de datos que existen y la calidad de los datos.

Los **sistemas de bases de datos** están formados por tres elementos fundamentales:

- La información que queremos guardar, es decir, los datos.
- El software que nos permite gestionar esta información. Este software lo denominamos sistema gestor de bases de datos (SGBD) o DBMS, **database management system**, de sus siglas en inglés.
- Los usuarios que pueden acceder a la información y los correspondientes derechos de acceso.

En el mercado encontramos un amplio catálogo de sistemas de bases de datos. Podemos agruparlos en dos grupos: las **BD relacionales** y las **BD no relacionales**.

Las BD relacionales organizan la información en tablas y en relaciones entre esas tablas. En este contexto **relación** y **tabla** son términos equivalentes, de ahí el nombre de este tipo de BD. Estos sistemas son útiles cuando tenemos datos estructurados y homogéneos. Además, las BD relaciones están diseñadas para asegurar cierta calidad en los datos. Esta característica las puede penalizar a la hora de procesar grandes cantidades de información, si la base de datos no está bien diseñada.

En cambio, las BD no relacionales son útiles para procesar datos poco, o nada, estructurados o bien, poco homogéneos. Algunas de estas bases de datos están diseñadas para representar estructuras jerárquicas. Por ejemplo, árboles semánticos o esquemas XML. Otras están diseñadas para representar grafos. Por ejemplo, redes sociales o referencias bibliográficas. Estos sistemas se han diseñado para procesar grandes cantidades de datos, pero se ha tenido que pagar un precio por ello. En ocasiones no podremos asegurar la consistencia de los datos. En breve explicaremos este concepto.

A continuación, podemos ver algunos ejemplos de los sistemas gestores más populares. En primer lugar, vemos ejemplos de sistemas gestores relacionales. Los primeros son sistemas con licencias mixtas comerciales y de software libre. Son muy populares porque funcionan muy bien en un conjunto amplio de aplicaciones. No obstante, empiezan a tener problemas si tienen que procesar muchos datos. Los segundos que vemos son sistemas comerciales por los que hay que pagar licencia para su uso. Son sistemas muy potentes pero que requieren de conocimientos avanzados para ser configurados. Son

los sistemas usados hasta ahora por las grandes corporaciones y empresas. Por último, vemos los sistemas gestores de bases de datos no relacionales. Se comercializan bajo diferentes tipos de licencias desde las de software libre hasta comerciales. Son sistemas diseñados para gestionar **Bigdata** y son relativamente fácil de configurar.

Sistemas gestores de bases de datos

Los sistemas gestores de bases de datos son los programas que permiten gestionar los datos, de un modo integral. No solo nos permiten acceder a toda la información almacenada en una BD de un modo eficaz, sino que nos permiten definir, y configurar, un conjunto de parámetros que nos permitirán asegurar la **calidad de los datos**.

La calidad de los datos es un requerimiento esencial en cualquier sistema de bases de datos. Los criterios que usamos para evaluar la calidad de un sistema de bases de datos son los siguientes:

- Seguridad
- Disponibilidad
- Confiabilidad
- Durabilidad

La **seguridad** de los datos hace referencia a los derechos de acceso y manipulación de los mismos. Estos derechos de acceso no sólo son importantes por una cuestión legal de derechos de autor y otro tipo de licencias, sino también, por la necesidad de tener un control del acceso de todos los usuarios que pueden, en un momento dado, modificar la información de la BD.

La **disponibilidad** de los datos hace referencia a todos aquellos aspectos, técnicos en su mayoría, que nos permitirán acceder a los datos en cualquier momento y bajo cualquier circunstancia. En función del tipo de disponibilidad que necesitemos, será necesario definir unas u otras políticas de replicación de datos.

La **confiabilidad** de la información es uno de los criterios más críticos en cualquier sistema de bases de datos. Si no podemos confiar en los datos con los que trabajamos, no podremos crear cualquier resultado que emana de ellos y, por tanto, resultarán del todo inútiles. Hay muchos aspectos que pueden, en un momento dado, condicionar la fiabilidad de los datos. Algunos de estos aspectos tendrán que ver con cuestiones técnicas, en la que poca cosa podremos hacer. Por ejemplo, después de una caída del sistema donde eventualmente se pueda haber perdido información (o simplemente no estamos seguros de si se ha perdido información o no). No obstante, en muchas ocasiones los problemas de confianza en los datos vienen de un diseño deficiente de la base de datos. Es sobre este aspecto, el diseño de bases de datos que incidiremos en cursos más avanzados de bases de datos.

Por último, la **durabilidad** de los datos hace referencia a todos aquellos aspectos que nos permitirán almacenar de un modo **permanente** la información de la base de datos.

En este caso estamos hablando de cuestiones esencialmente técnicas, y serán, los administradores de las bases de datos los responsables de asegurarla.

Desde un punto de vista técnico, decimos que un sistema de bases de datos asegura la calidad de los datos si las transacciones satisfacen las **propiedades ACID**.

Definamos primero que es una transacción. Una **transacción** es el conjunto de operaciones de una base de datos con sentido lógico. El ejemplo típico de transacción es el de la transferencia bancaria. Cuando un usuario hace una transferencia de una cuenta bancaria a otra se tiene que modificar el saldo de las dos cuentas. Si en el tiempo de hacer las dos operaciones ocurre algún tipo de incidencia y no se puede hacer la 2ª operación, cualquiera que sea de las dos modificaciones de saldo, se querrá que se deshagan todas las acciones de la transacción y el saldo en las dos cuentas de modo que el saldo de las cuentas vuelva a ser el que había antes de iniciar la transferencia.

Esta primera propiedad que tiene que satisfacer la transacción se llama **atomicidad** y es la primera de las propiedades ACID. Una transacción es atómica. Es decir, o se realizan todas las operaciones de la transacción o ninguna. Siguiendo el ejemplo, o se modifican los dos saldos, o ninguno.

La segunda propiedad es la **consistencia** y está íntimamente ligada con el criterio de confiabilidad que hemos explicado hace unos minutos. Los datos de una BD tienen que ser válidos. Consistente y válido, en este contexto, son sinónimos y se refiere al hecho que los valores que se encuentren en una BD tienen que ser correctos. Siguiendo el ejemplo de la transferencia bancaria, el valor que restamos en una cuenta tiene que ser el mismo que sumamos en la otra cuenta. En caso que ocurra algún tipo de incidente mientras no se haya confirmado el final de la transacción, será el sistema gestor de la base de datos el encargado de restaurar la BD al último estado consistente.

La tercera propiedad que tiene que satisfacer un sistema gestor de base de datos es la de **aislamiento** (Isolation en inglés). Múltiples usuarios pueden acceder a los mismos datos e intentar modificarlos simultáneamente. Un SGBD tiene que ser capaz de aislar las transacciones, de modo que no se interfieran las unas a las otras, ordenar su ejecución y asegurar que se ejecuten correctamente.

Por último, la propiedad de **durabilidad** coincide con el criterio con el mismo nombre. El SGBD tiene que asegurar, mediante copias de seguridad y técnicas de recuperación, que cuando una transacción ha concluido satisfactoriamente su ejecución y su resultado es permanente.

Conclusión

En resumen, en cualquier sistema de bases de datos es importante asegurar la calidad de los datos. Las bases de datos relacionales satisfacen un conjunto de propiedades que ayudan a asegurar esa calidad. En cambio, las bases de datos no relaciones no siempre

las pueden asegurar. En cualquier caso, siempre es necesario un buen diseño de la base de datos y usar el tipo de BD, relacional o no relacional, que mejor se adapte a los datos a manipular. En próximos cursos profundizaremos sobre los aspectos de diseño de bases de datos. Hasta la próxima.

Bibliografía de referencia

- *Database System Concepts Sixth Edition*, Avi Silberschatz, Henry F. Korth, S. Sudarshan, McGraw-Hill, ISBN 0-07-352332-1.
- [Fundamentos de sistemas de bases de datos](#), Elmasri, Ramez, Madrid [etc.] : Pearson Addison Wesley, cop. 2007, 5ª ed.
- *Introducción a los sistemas de Bases de Datos*, C.J. Date, Vol.1, 7ª edición, Prentice Hall, 2001.

Autor: Oriol Ramos
Departamento de Ciencias de la
Computación. UAB

