

Big Data

Big data es un concepto que se refiere a grandes volúmenes de datos que son muy variados y veloces, al punto de que resulta muy complicado capturarlos y procesarlos con métodos tradicionales. Para su uso es necesaria una herramienta cliente/servidor para la gestión de Bases de Datos, diseñada para que las organizaciones puedan controlar y gestionar grandes volúmenes de contenidos no estructurados en un único repositorio con el objetivo de reducir los costes y los riesgos asociados a la pérdida de información. A continuación, se definirán varios ejemplos de motores que se dedican a la gestión de bases de datos.

SQLite

SQLite es un sistema de gestión de bases de datos relacional contenida en una librería de C y que, a diferencia de los motores convencionales, se incorpora directamente en el programa desarrollado sin la necesidad de usar un sistema de cliente-servidor.

Generalmente se emplea la sintaxis de PostgreSQL usando tipos de datos dinámicos y débilmente tipados, lo que no garantiza la integridad de los datos.

Este suele ser la opción más popular para programas que se ejecutan únicamente en local o para el almacenamiento de los datos del cliente de un servicio en su propio sistema. Gracias a esa popularidad, SQLite actualmente está disponible para una gran variedad de lenguajes de programación.

Fabricante y Licencia

SQLite fue diseñado y desarrollado por Dwayne Richard Hipp mientras trabajaba para General Dynamics. SQLite es Open Source bajo una licencia de tipo GNU General Public License.

Modo de comercialización

El código de SQLite está disponible a través de Fossil, un sistema de control de versiones distribuido desarrollado también por Dwayne Richard Hipp.

Apache Hive

Apache Hive es un proyecto especializado en el almacenamiento de datos (data warehouse) creado sobre Apache Hadoop para proveerle funcionalidades de consulta y análisis de datos con una interfaz del estilo de SQL para realizar consultas en diversas bases de datos y sistemas de archivos que se integran con Hadoop.

A diferencia de las consultas SQL tradicionales, que deben ser implementadas en la API MapReduce de Java para poder ejecutar las aplicaciones y consultas sobre datos distribuidos, Hive provee la abstracción necesaria de SQL

Fabricante y Licencia

Apache Hive fue inicialmente desarrollado por Facebook, pero actualmente es usado y desarrollado por otras compañías como Netflix o FINRA. Además, Amazon mantiene un fork del código, el cual tiene incluido en Amazon Elastic MapReduce, uno de los servicios ofrecidos dentro de los AWS. Apache Hive es Open Source bajo una licencia de tipo Apache License 2.0

Modo de comercialización

Apache Hive provee el software a través de su página web, en la cual se pueden encontrar desde la documentación hasta un historial con los cambios realizados en cada versión del sistema.

Microsoft SQL Server

Microsoft SQL Server es un sistema de gestión de bases de datos que soporta transacciones y procedimientos almacenados. También incluye un entorno gráfico de administración, que permite el uso de comandos DDL y DML gráficamente. Además, permite trabajar en modo cliente-servidor, donde la información y datos se alojan en el servidor y los terminales o clientes de la red solo acceden a la información.

Además, permite administrar información de otros servidores de datos haciéndolo más versátil frente a otros motores de bases de datos. Por otro lado, este sistema incluye una versión reducida, llamada MSDE con el mismo motor de base de datos, pero orientado a proyectos más pequeños.

Fabricante y Licencia

Microsoft SQL Server es un sistema de gestión de base de datos relacional, desarrollado por la empresa Microsoft. Es un software propietario bajo la licencia Microsoft EULA. Mediante esta licencia, que es de uso y no de compra de producto, Microsoft se reserva todos los otros derechos para sí mismos. Con esta licencia no tienes derechos a:

- Trabajar sobre las limitaciones técnicas del software.
- Realizar ingeniería inversa, descompilar o desensamblar el software.
- Eliminar, minimizar, bloquear o modificar cualquier noticia de Microsoft o sus proveedores.
- Usar el software para uso comercial o que generen ingresos.
- Usar el software en cualquier forma en contra de la ley o para crear y/o propagar Malware.
- Compartir, publicar, distribuir o prestar el software.

Modos de comercialización

Microsoft SQL Server se vende como licencia. Sin embargo, los usuarios que quieran utilizar SQL Server, siendo todo sobre el programa, como solicitar y almacenar información, necesitan una CAL (Client Access License).

- **Modelo de servidor + CAL:** Este modelo de licencia es una opción más económica, se compra con la licencia y se tiene que comprar CAL adicionales para usuarios o dispositivos.
- **Licencias por núcleo:** Al elegir la concesión de licencias por núcleo, concede la licencia a varios núcleos del procesador físico del servidor. Con Microsoft SQL Server 2019, al menos 4 núcleos siempre deben tener licencia. Este modelo de licencia suele ser elegido por grandes empresas.
- **Client Access License (CAL)** (Una licencia de acceso de cliente) Una licencia de acceso de cliente está destinada a proporcionar derechos de uso para permitir que los usuarios o dispositivos accedan al software del servidor. Esto significa que cada computadora o usuario que necesita acceso al software del servidor necesita una CAL. Existen dos tipos de CAL.
- **CAL de usuario** da acceso a la función antes mencionada a un usuario en la red. Esto significa que este usuario también puede acceder a la red del servidor a través de otros dispositivos.
- **CAL de dispositivo** da acceso a un dispositivo. Esto significa que todos los usuarios de este dispositivo pueden iniciar sesión para acceder al servidor. A menudo, esto solo es beneficioso cuando hay más usuarios que dispositivos dentro de una organización,¹³

Versiones

Cada versión de SQL Server posee distintas versiones con distintos precios (para cada versión) que dependen también en la configuración física del servidor. A continuación, se presentan las versiones principales:

- **Enterprise:** Contempla todas las características (deshabilitadas en otras ediciones). Es el tipo de versión con más privilegios existente en el mercado.
- **Developer:** Una edición con las mismas características que la Enterprise, con el fin de ser instalada solamente en ambiente de desarrollo y no en producción. Si se desarrolla para una edición Standard hay que tener en cuenta las características deshabilitadas para esta versión.
- **Standard:** Una versión limitada según la configuración del servidor y sus características, diseñada para servidores inferiores.
- **Express:** Una versión gratuita que posibilita la creación de bases de datos limitadas con características básicas, con el fin de apoyar aplicaciones que necesiten una solución simple para almacenamiento de una cantidad limitada de datos, o usuarios que sus recursos y necesidades son limitados.
- **SQL Azure:** Es una versión de SQL Server en la nube, que permite pagar mensualmente por el servicio sin la necesidad de mantener un servidor.

ORACLE

Oracle Corporation es una compañía especializada en el desarrollo de soluciones de nube y locales. En cuanto a las características principales de Oracle, este contiene:

- Modelo relacional: los usuarios visualizan los datos en tablas con el formato filas/columnas.
- Herramienta de administración gráfica intuitiva y cómoda de utilizar.
- Lenguaje de diseño de bases de datos muy completo (PL/SQL)
- Gestión de usuarios: agilidad en los trámites, reducción de costes y seguridad en el control de las personas que acceden a las aplicaciones y a los sistemas.
- Alta disponibilidad: escalabilidad, protección y alto rendimiento para la actividad empresarial.
- Protección de datos: seguridad completa en el entorno de producción y de pruebas y gestión de copias de seguridad.

Fabricante y Licencia

Fabricado por Oracle Corporation que es una de las mayores compañías de software del mundo, cuyas oficinas centrales se localizan en EEUU. Cuenta además con herramientas propias de desarrollo para realizar potentes aplicaciones, como Oracle Designer.

Oracle es un software propietario y además es conocido por su lucha contra el software libre, creando incluso un informe dirigido al gobierno de los Estados Unidos aconsejando el no usar el software de código abierto.

Modo de comercialización:

En cuanto a la comercialización de productos de software Oracle, cabe destacar que es la compañía líder en tecnología Internet, en la actualidad el 90% de las empresas que operan en comercio electrónico a través de Internet utilizan los productos de software Oracle para sus operaciones.

MongoDB

MongoDB (del inglés humongous, "enorme") es un sistema de base de datos NoSQL, orientado a documentos y de código abierto. En lugar de guardar los datos en tablas, tal y como se hace en las bases de datos relacionales, MongoDB guarda estructuras de datos BSON (una especificación similar a JSON) con un esquema dinámico, haciendo que la integración de los datos en ciertas aplicaciones sea más fácil y rápida.

- **Consultas ad hoc:** MongoDB soporta la búsqueda por campos, consultas de rangos y expresiones regulares. Las consultas pueden devolver un campo específico del documento, pero también puede ser una función definida por el usuario para su mejor ocupación.
- **Indexación:** Cualquier campo en un documento de MongoDB puede ser indexado, al igual que es posible hacer índices secundarios. El concepto de índices en MongoDB es similar al empleado en base de datos relacionales.
- **Replicación:** MongoDB soporta el tipo de replicación primario-secundario. Cada grupo de primario y sus secundarios se denomina replica set. El primario puede

ejecutar comandos de lectura y escritura. Los secundarios replican los datos del primario y sólo se pueden usar para lectura o para copia de seguridad, pero no se pueden realizar escrituras. Los secundarios tienen la habilidad de poder elegir un nuevo primario en caso de que el primario actual deje de responder.

- **Balanceo de carga:** MongoDB puede escalar de forma horizontal usando el concepto de shard. El desarrollador elige una clave de sharding, la cual determina cómo serán distribuidos los datos de una colección. Los datos son divididos en rangos (basado en la clave de sharding) y distribuidos a través de múltiples shard. Cada shard puede ser una réplica set. MongoDB tiene la capacidad de ejecutarse en múltiples servidores, balanceando la carga y/o replicando los datos para poder mantener el sistema funcionando en caso que exista un fallo de hardware. La configuración automática es fácil de implementar bajo MongoDB y se pueden agregar nuevos servidores a MongoDB con el sistema de base de datos funcionando.
- **Almacenamiento de archivos:** MongoDB puede ser utilizado como un sistema de archivos, aprovechando la capacidad de MongoDB para el balanceo de carga y la replicación de datos en múltiples servidores. Esta funcionalidad, llamada GridFS e incluida en la distribución oficial, implementa sobre los drivers, no sobre el servidor, una serie de funciones y métodos para manipular archivos y contenido. En un sistema con múltiples servidores, los archivos pueden ser distribuidos y replicados entre los mismos de forma transparente, creando así un sistema eficiente tolerante de fallos y con balanceo de carga.
- **Agregación:** MongoDB proporciona un framework de agregación que permite realizar operaciones similares al "GROUP BY" de SQL. El framework de agregación está construido como un pipeline en el que los datos van pasando a través de diferentes etapas en las cuales estos datos son modificados, agregados, filtrados y formateados hasta obtener el resultado deseado. Todo este procesamiento es capaz de utilizar índices si existieran y se produce en memoria. Asimismo, MongoDB proporciona una función MapReduce que puede ser utilizada para el procesamiento por lotes de datos y operaciones de agregación.
- **Ejecución de JavaScript del lado del servidor:** MongoDB tiene la capacidad de realizar consultas utilizando JavaScript, haciendo que estas sean enviadas directamente a la base de datos para ser ejecutadas.

Fabricante

El desarrollo de MongoDB comenzó en 2007 de la mano de 10gen Inc. (ahora llamada MongoDB Inc.) cuando desarrollaban una plataforma como servicio (PaaS) similar al conocido Google App Engine. En 2009 MongoDB fue lanzado como un producto independiente y publicado bajo la licencia de código abierto AGPL.

Modo de comercialización:

MongoDB se puede obtener de forma gratuita bajo la licencia pública general de Affero (AGPL) de GNU. Los drivers para los lenguajes de programación están bajo la licencia de Apache. Además de estas licencias, MongoDB Inc. ofrece una licencia comercial que incluye características avanzadas (como integración con SASL, LDAP, Kerberos, Simple Network Management Protocol y búsqueda de texto de Rosette Linguistics Platform de Basis Technology), herramientas de gestión, monitorización y backup, así como soporte.

Conclusiones

Algunos de los aspectos aprendidos y que de gran peso es la base de datos su definición, requerimiento, ventajas y características donde podemos decir que la base de datos es una colección de datos o información usados para dar servicios a muchas aplicaciones al mismo tiempo. Existen muchas empresas dedicadas exclusivamente a la creación, gestión y mantenimiento de bases de datos ya que hoy en día los datos tienen un inmenso valor para las empresas. Con ellos, pueden predecir cual será la mejor forma o el mejor público al que vender su producto. Además, se pueden usar para hacer mas atractivo el uso de un software o conocer que tipo de persona requeriría cierto producto y en qué circunstancias.

En resumen, concluimos que las bases de datos son una parte esencial del software y que una buena base de datos puede generar grandes tanto beneficios en la producción y gestión de un software, como en un producto.

Jorge Acevedo de León

Javier Martín de León

Adrián Epifanio Rodríguez Hernández

Sergio Tabares Hernández