

Trabajo Práctico 6 - BASES DE DATOS 2

Hayk Kocharyan
757715@unizar.es

Pedro Tamargo Allué
758267@unizar.es

Jesús Villacampa Sagaste
755739@unizar.es

Juan José Tambo Tambo
755742@unizar.es

May 2, 2020

Contents

1	Enunciado	2
2	SGBD	3
2.1	Oracle	3
2.2	PostgreSQL	4
2.3	DB2	4
2.4	Access	5
2.5	VoltDB	6
2.6	Apache Cassandra	7
2.7	MongoDB	7
2.8	Apache HBase	8
3	SGBD	9
3.1	Oracle	9
3.2	MySQL	9
3.3	PostgreSQL	9
3.4	DB2	9
3.5	Microsoft SQL Server	9
3.6	Microsoft Access	9
3.7	Caché	10
3.8	VoltDB	10
3.9	Cassandra	10
3.10	MongoDB	10
3.11	HBase	10
	Referencias	11

1 Enunciado

Para la realización del trabajo se ha elegido el problema del *Gobierno de Aragón*. El problema proporcionado en el enunciado dice lo siguiente:

“El Gobierno de Aragón quiere gestionar información turística de la región para ofrecerla a través de un sitio web donde los usuarios podrán hacer comentarios y recomendaciones.”

Por lo tanto la expansión del enunciado es la siguiente:

Se pretende desarrollar una base de datos para almacenar la información turística de la región de Aragón. Se llevará acabo a través de un sitio web mediante la cual, los usuarios puedan consultar y realizar recomendaciones de puntos turísticos, así como comentar otras. También, este sistema deberá permitir almacenar usuarios. Ya que la página web permitirá realizar publicaciones, recomendaciones y comentarios se almacenarán en la base de datos. Se deberá garantizar el acceso concurrente a la base de datos debido a que múltiples usuarios pueden realizar publicaciones y comentarios de manera simultánea. Respecto a la garantía de la calidad del servicio primaremos la consistencia de los datos frente a la disponibilidad de la misma.

En relación a la estimación del trafico web, será de unos **10.000** usuarios a la semana, **15.000** en alta temporada.

Por lo mencionado anteriormente, este sistema gestor deberá garantizar una buena interoperabilidad con los lenguajes de programación utilizados en este proyecto.

Si fuese necesario se podría utilizar la infraestructura disponible en las ciudades de Huesca y Zaragoza.

Por ello, en base a estos requisitos buscaremos un sistema gestor *OLTP (Online Transaction Processing)* ya que un gestor OLAP está enfocado para en el análisis. Respecto al volumen de datos, el sistema no se prevee que almacene grandes volúmenes de datos, por lo tanto la escalabilidad no es un factor primordial.

Según los datos del presupuesto de turismo proporcionados por el presupuesto de la comunidad de Aragón del año 2020 **insertar referencia aquí**, turismo cuenta con una cantidad de 13.100 € (trece mil cien euros).

Con estas condiciones suponemos que vamos a contar con un presupuesto similar.

Dado que el Gobierno de Aragón es una entidad pública se buscará el uso de un SGBD consolidado, y que ofrezca buen soporte en el tiempo.

IMPORTANTE

- SO soportado
- Soporte y popularidad del producto
- Adecuación al dominio requerido (a que está orientado y eso)
- Presupuesto, Costes, Licencias...
- Escalabilidad
- Soporte para concurrencia
- CAP
- Experiencia previa del usuario **para la conclusión**
- Disponibilidad de herramientas para soporte (consultas, admin, copias..)
- Calidad del soporte ante fallos y problemas
- Casos similares de nuestro ámbito
- Gestores ya en uso **para la conclusión**
- Opiniones de otros usuarios <https://www.trustradius.com/> **pagina para mirar opiniones**
- Mismo vendedor aka reducción de costes **para conclusión**

2 SGBD

2.1 Oracle

Oracle es un sistema gestor de bases de datos multimodelo (soporta modelo relacional y modelo objeto relacional), perteneciente a la compañía *Oracle Corporation* [78]. En 1977 fue fundada la compañía *SDL (Software Development Laboratories)*, este *SGBD* proviene de un proyecto financiado por la *CIA* para diseñar un sistema especial de bases de datos con código clave “*Oracle*” [77]. Los fundadores de *SDL* habían leído un artículo en la revista *IBM research* donde se describía una versión preliminar de *SQL* basado en el artículo de *E. F. Codd*¹, donde propone el *modelo relacional*. En 1984, *SDL* adopta el nombre de *Oracle Corporation*.

En su versión 19, la instalación de este *SGBD* se puede realizar *on-premise* (en los servidores de la empresa) o en el *cloud* de *Oracle*. Si elegimos la opción *on-premise*, los sistemas operativos en los que es posible instalar este *SGBD* son: *Windows*, *Linux* y *Solaris* [45].

Oracle es una de las bases de datos líderes del mercado en bases de datos operacionales [25]. Cuenta con mucha documentación disponible desde su página web² y además cuenta con una comunidad muy activa en la resolución de incidencias y dudas en páginas como *StackOverflow* [27].

El dominio de aplicación de este *SGBD* se encuentra desde el desarrollo de aplicaciones hasta servicios de almacenes de datos [44]. **Algo sobre el ámbito principal? No sé cual es y nadie dice nada.**

El sistema de licenciamiento de este *SGBD* depende del número de usuarios y el número de procesadores que tenga el servidor donde se ha de instalar. Para cada “*Processor License*” se aplica el “*Core factor*”³ y ese es el número de “*Processor Licenses*” necesarios. Para el cálculo de “*Named User Licenses*” se aplicará el máximo entre:

- $ProcessorLicenses * CoreFactor * NumUserMinimun$
- $NumUsuarios + NumDispositivos$

Donde, *NumUserMinimun* se corresponde con el número de usuarios mínimos por cada “*Processor License*” [47]. Los precios de “*Processor License*” y “*Named User Licenses*” en la versión *Enterprise Edition* se corresponden con: 47500\$ y 950\$ [48]. El resto de funcionalidades no añadidas, como las de almacenes de datos, se adquieren por separado.

En el tema de escalabilidad, *Oracle* cuenta con *RAC (Real Application Cluster)*, una tecnología que permite escalar una base de datos a un cluster de servidores ejecutando multiples instancias sobre una misma base de datos de manera transparente al usuario, aprovechando los recursos del clúster. [46]

El control de la concurrencia en este *SGBD* se realiza mediante los mecanismos de bloqueo del mismo para garantizar la consistencia de los datos entre transacciones. *Oracle* se encarga de gestionar los mecanismos de bloqueo automáticamente, de tal forma que no es necesario que el usuario escriba el bloqueo de forma explícita.[43] Los mecanismos de bloqueo utilizados por este *SGBD* se dividen en dos grupos, bloqueos exclusivos y bloqueos compartidos. [2]

Para el desarrollo con este *SGBD* se puede utilizar el driver *ODBC* (o *JDBC* si se usa *Java*) desde la propia página de la compañía.

El uso de herramientas de soporte complementarias como puede ser *DBeaver*⁴ y otras opciones de apoyo al desarrollo es posible configurando (si fuera necesario) el driver de conexión a la base de datos.

Las copias de seguridad se pueden realizar utilizando la herramienta *RMAN*. Esta herramienta es específica para las bases de datos *Oracle* y permite realizar copias de seguridad sobre las bases de datos y realizar una vuelta a una versión estable si disponemos de una copia de seguridad [42].

¹https://es.wikipedia.org/wiki/Edgar_Frank_Codd

²<https://docs.oracle.com/en/database/oracle/oracle-database/19/lnpls/index.html>

³<http://www.oracle.com/us/corporate/contracts/processor-core-factor-table-070634.pdf>

⁴<https://dbeaver.io/>

El soporte técnico de este *SGBD* se adquiere por separado, y se relaciona con cada uno de los productos adquiridos. [48] **COMPROBAR ESTO**

2.2 PostgreSQL

PostgreSQL, también llamado *Postgres*, es un sistema de gestión de bases de datos relacional orientado a objetos y de código abierto. Como muchos otros proyectos de código abierto, el desarrollo de *PostgreSQL* no es manejado por una empresa o persona, sino que es dirigido por una comunidad de desarrolladores que trabajan de forma desinteresada, altruista, libre o apoyados por organizaciones comerciales. Dicha comunidad es denominada *PGDG*⁵. [66]

PostgreSQL está publicado bajo *PostgreSQL License*, una licencia *Open Source*, similar a *BSD*⁶ o *MIT*⁷. [50]

Uno de los puntos fuertes de *Postgres* es el cumplimiento de las propiedades ACID⁸ y su gestión de la concurrencia. Aseguran que una lectura nunca bloquea escrituras y viceversa. Esto se consigue a través de un mecanismo llamado *MVCC*⁹. Cada transacción en *Postgres* obtiene una ID de transacción llamada *XID*. Esto incluye transacciones de una sola declaración, como insertar, actualizar o eliminar, así como envolver explícitamente un grupo de declaraciones juntas a través de *BEGIN-COMMIT*. Cuando comienza una transacción, *Postgres* incrementa un *XID* y lo asigna a la transacción actual. [15]

Independientemente de la plataforma y la arquitectura que usemos, *PostgreSQL* está disponible para los diferentes SO, *Unix*, *Linux* y *Windows*, en 32 y 64 bits. Ésto hace de *PostgreSQL* un sistema multiplataforma y también hace que sea más rentable con instalaciones a gran escala. Además tiene más de 20 años de desarrollo activo y en constante mejora. No se han presentado nunca caídas de la base de datos. Ésto es debido a su capacidad de establecer un entorno de alta disponibilidad gracias a *Hot-Standby*¹⁰. [52]

Postgres está diseñado para ambientes de alto volumen, lo que hace que la velocidad de respuesta en inserciones o actualizaciones pueda parecer lenta en comparación con bases de datos de pequeño tamaño. [54]

Otro aspecto muy importante a tener en cuenta a la hora de elegir *SGBD* es el soporte que proporcionan. En el caso de *PostgreSQL* no cuenta con soporte en línea o telefónico. Sin embargo, cuenta con foros oficiales donde usuarios exponen sus dudas y son los propios usuarios de la comunidad los que responden. Cabe resaltar que esta comunidad de usuarios es una de las más activas del mercado. Además, también ofrece soporte empresarial como *EnterpriseDB*¹¹ o *TodoPostgreSQL*¹². [54]

Postgres es uno de los gestores con más soporte y recorrido en el mercado, y por tanto, cuenta con una herramienta administración llamada *pgAdmin* (desde 2016 en la versión *pgAdmin4*), que facilita la gestión y administración de bases de datos ya sea mediante instrucciones SQL o con ayuda de un entorno gráfico. Permite acceder a todas las funcionalidades de la base de datos; consulta, manipulación y gestión de datos, incluso opciones avanzadas como manipulación del motor de replicación *Slony-I*¹³. A parte de este entorno de escritorio, también existen otras herramientas como *psql* (cliente de consola) o *PhpPgAdmin* (entorno web). [49]

2.3 DB2

IBM DB2 es un gestor de bases de datos relacional orientada a empresas capaz de gestionar datos estructurados y no estructurados en entornos públicos, privados y locales. [12] Actualmente es una marca comercial, propiedad

⁵PostgreSQL Global Development Group

⁶https://es.wikipedia.org/wiki/Licencia_BSD

⁷https://es.wikipedia.org/wiki/Licencia_MIT

⁸Atomicity, Consistency, Isolation y Durability (Atomicidad, Consistencia, Aislamiento y Durabilidad en castellano)

⁹Multi Version Concurrency Control

¹⁰Hot-Standby nos permite que los clientes puedan realizar consultas de solo lectura mientras que los servidores están en modo de recuperación o espera. Así podemos hacer tareas de mantenimiento o recuperación sin bloquear completamente el sistema [53]

¹¹<https://www.enterprisedb.com>

¹²<https://todopostgresql.com/>

¹³Slony-I es un sistema asíncrono de replicación maestro-esclavo para PostgreSQL, que proporciona soporte para conexión en cascada y conmutación por error. Asíncrono significa que cuando una transacción de base de datos se ha comprometido con el servidor maestro, todavía no se garantiza que esté disponible en esclavos. La conexión en cascada significa que las réplicas se pueden crear (y actualizar) a través de otras réplicas, es decir, no necesitan conectarse directamente al maestro. [56]

de *IBM*, aunque posee una versión gratuita, *DB2 Express-C*[10], que permite desarrollar aplicaciones con las funcionalidades básicas de *DB2*.

Entre sus principales características destacan la capacidad de manejar objetos de hasta 2 GB, definición de funciones y datos/tipos por parte del usuario, mecanismos para garantizar la integridad referencial, *SQL* recursivo¹⁴, tratamiento de archivos multimedia, procesamiento paralelo, commit en dos fases y *backup* on-line y offline. Además permite la conexión directa de una aplicación escrita en *Java* con la base de datos utilizando *JDBC*, pudiendo ser una base local o remota.[13]

El gestor utiliza *prueXML*¹⁵ para almacenar documentos en formato *XML*, para realizar operaciones y búsquedas dentro del mismo e integrarlo con búsquedas relacionales.[59] De esta manera se puede realizar consultas de tipos de datos no tradicionales. Además utiliza un sistema multiprocesador *SMP* simétrico y un sistema de procesador paralelo masivo, lo que permite su alto rendimiento.

Se trata de un sistema altamente escalable al dar soporte a volúmenes de datos de *petabytes* y acelerar las consultas, pudiendo escalar desde un *laptop* a *clusters* de servidores empresariales paralelos.[9]. Asimismo, *DB2* ofrece múltiples opciones de despliegue (híbrido, *on premise*, cloud), lo que da soporte a un amplio rango de cargas de trabajo.

DB2 utiliza un sistema de *commit-rollback* diferente a otros gestores. Todos cambios en la base se almacenan en el *Log Buffer*, el cual escribe en disco cuando se realiza un *commit*. Un *commit* no es correcto hasta que se ha copiado toda información en disco.[14]

Las plataformas que admite el gestor son *AIX*, *HP-UX*, *UNIX*, *Linux*, *Solaris* y *Windows*, además de poseer una versión para *Z/OS*¹⁶, la cual está optimizada para *SOA*, *CRM* y almacenamiento de datos.[8]

Por último, el gestor ofrece una serie de herramientas de administración, gestión de rendimiento, recuperación, réplicas y aplicaciones. La mayoría de estas herramientas proporcionan una interfaz gráfica de usuario (*GUI*) y una interfaz *ISPF*¹⁷ que permiten al usuario realizar la mayoría de tareas de *DB2* de forma interactiva.[11]

2.4 Access

Microsoft Access es un sistema de gestión de bases de datos incluido en el paquete ofimático denominado *Microsoft Office*, sucesor de *Embedded Basic*. *Access* es un gestor de datos que utiliza los conceptos de bases de datos relacionales y pueden manejarse por medio de consultas e informes. Está adaptado para recopilar datos de otras utilidades como *Excel* o *SharePoint*. A partir del 22 de septiembre de 2015 la versión más nueva es *Access 2016*. [55]

Los requisitos para la utilización de este gestor son un sistema operativo *Windows 10*, un procesador de 2 núcleos a 1'6GHz, una memoria RAM de 2GB, un espacio libre en disco duro de 4 GB y una cuenta Microsoft [31]. El precio para adquirir una licencia para 1 PC de *Access* es de 135€. [30]

Las bases de datos *access* son útiles para pequeñas y medianas empresas o departamentos individuales, ya que cuentan con una capacidad limitada. Cualquier sector cuyo uso vaya más allá de las 2 GB descubrirá sus limitaciones [26]. Para evitar esta limitación de tamaño, se puede crear vínculos a tablas de otras bases de datos de *Access*. Puede crear vínculos a tablas de varios archivos de base de datos, cada uno de los cuales puede tener hasta 2 gigabytes de tamaño. [29]

Microsoft Access ofrece consultas parametrizadas¹⁸. Estas consultas y tablas *Access* se pueden referenciar desde otros programas como *VB6*¹⁹ y *.NET* a través de *DAO* o *ADO*. [51]

¹⁴<https://geeks.ms/jirigoyen/2009/05/22/recursividad-con-sql-server/>

¹⁵Función de almacenamiento XML nativo.

¹⁶Sistema actual de las computadoras de *IBM*

¹⁷Interactive System Productivity Facility

¹⁸Se denominan consultas paramétricas a aquellas consultas, normalmente de selección, que bajo un campo por el que se desea establecer un criterio de filtrado, en vez de implicar a valores constantes que hacen que dicha consulta al ser ejecutada siempre realice "lo mismo", se implique a campos (especificados entre corchetes []) a los cuales se les asignan valores diferentes para cada ejecución de la consulta (normalmente serán solicitados al usuario mediante una ventana en pantalla).[1]

¹⁹Visual Basic 6.0

También está el formato de base de datos *Jet* ²⁰, que puede contener la aplicación y los datos en un archivo. Esto hace muy conveniente para distribuir la aplicación completa a otro usuario, que puede ejecutarla en entornos desconectados. [51]

El límite técnico de las bases de datos *access* es de 255 usuarios al mismo tiempo, pero el límite real es entre 10 y 80 usuarios dependiendo del tipo aplicación. [26]

2.5 VoltDB

VoltDB es una base de datos en memoria diseñada por Michael Stonebraker, Sam Madden y Daniel Abadi. Es un *RDBMS* ²¹ compatible con ACID que utiliza una arquitectura de nada compartido ²². Incluye tanto ediciones empresariales como comunitarias. La edición comunitaria está licenciada bajo la *GNU Affero General Public License*. [67]

VoltDB es una base de datos relacional *NewSQL* ²³ que admite acceso *SQL* desde procedimientos almacenados Java precompilados. La unidad de transacción es el procedimiento almacenado, escrito en Java intercalado con *SQL*. *VoltDB* se basa en la partición horizontal hasta el subproceso de hardware individual para escalar, k-safety (replicación síncrona) para proporcionar una alta disponibilidad y una combinación de instantáneas continuas y registro de comandos para una mayor durabilidad (recuperación de fallos). [67]

Una gran desventaja de este *SGBD* es que solo se ejecuta en *Linux* y *Mac*, y por tanto para su utilización en *Windows* haría falta un despliegue en *contenedores docker* o la utilización de máquinas virtuales. [3]

El punto clave de *VoltDB* es su velocidad. Puede procesar hasta 45 veces mas transacciones por segundo que otros gestores tradicionales como *MySQL*, *Oracle* y *PostgreSQL*. *VoltDB* procesó 53.000 transacciones por segundo (TPS) frente a las 1.155 de otros *DBMS* sobre hardware idéntico. [4] Debido a esto, se debe ejecutar en servidores ricos en memoria multi-core y aunque esté disponible de manera gratuita bajo la licencia *GPL* ofrecen un precio de suscripción anual es de 15 dólares por año para una configuración de cuatro servidores.

VoltDB permite operacionalizar *modelos ML*. *VoltDB* convierte automáticamente el *modelo ML* en un proceso ejecutable como una función definida por el usuario (UDF), que puede implementarse en un entorno de producción. El modelo integrado ingiere / entrena continuamente datos nuevos / históricos para mantenerse actualizado y relevante. El ML en la base de datos le permite tomar decisiones inteligentes y procesables en tiempo real con una latencia muy baja en la transmisión de datos. [5]

VoltDB se ejecuta en un entorno distribuido en contenedores. Hasta ahora, orquestar bases de datos *SQL* en *Kubernetes* ²⁴ había sido muy desafiante, ya que los sistemas de bases de datos operacionales tienen estado y no pueden ser acelerados o reducidos en cualquier momento. Los desarrollos recientes aseguran la orquestación de contenedores con *Kubernetes*, convirtiendo muchas tareas tediosas y complejas en algo tan simple como un archivo de configuración declarativo, y permitiendo una implementación continua y frecuente. [5]

²⁰Jet es un motor de bases de datos que permitiría mantener los archivos .mdb a través de ODBC [64]

²¹Sistema de gestión de bases de datos relacionales

²²Una arquitectura de nada compartido es una arquitectura de computación distribuida en la que cada solicitud de actualización es satisfecha por un solo nodo (procesador / memoria / unidad de almacenamiento). La intención es eliminar la contención entre nodos. Los nodos no comparten (acceso independiente) memoria o almacenamiento [58]

²³NewSQL es una clase de sistemas de gestión de bases de datos relacionales que buscan proporcionar la escalabilidad de los sistemas NoSQL para cargas de trabajo de procesamiento de transacciones en línea (OLTP) mientras mantienen las garantías ACID de un sistema de base de datos tradicional. [65]

²⁴Kubernetes es un sistema de código libre para la automatización del despliegue, ajuste de escala y manejo de aplicaciones en contenedores que fue originalmente diseñado por Google y donado a la Cloud Native Computing Foundation. Soporta diferentes entornos para la ejecución de contenedores, incluido Docker. [60]

2.6 Apache Cassandra

Cassandra es un sistema gestor de bases de datos *NoSQL* (*Not Only SQL*), con almacenamiento basado en columnas. Fue creado desarrollado por *Avinash Lakshman*, uno de los autores de *Amazon's Dynamo*²⁵ por *Facebook* como un proyecto para mejorar la búsqueda en la bandeja de entrada de la plataforma. En 2008, *Facebook* liberó a *Cassandra* como un proyecto open-source en *Google Code*²⁶. En Marzo de 2009, *Apache* acogió este proyecto, transformándolo en un proyecto de *Apache Incubator* (proyecto en la que los proyectos open-source se pueden convertir en nuevos proyectos de alto nivel de la fundación *Apache* [17]). En Febrero de 2010 se graduó como un proyecto de alto nivel. [72]

Cassandra, al estar programado sobre *Java*, es multiplataforma, es decir, se puede instalar sobre cualquier sistema operativo siempre que cuente con *Java* (recomiendan la última versión de *Java* 8). La opción más habitual es utilizar un servidor con SO *Linux*. [23]

Este *SGBD* pertenece a *Apache Software Foundation*, una compañía sin ánimo de lucro que desde 1999 ha desarrollado software open-source. *Cassandra* se graduó como un proyecto de *Apache Incubator* en 2010, y desde entonces ha liberado 11 versiones. Sigue manteniendo de forma activa 3 de ellas. [22]

El dominio de aplicación de *Cassandra* se corresponde con el análisis de datos en entornos distribuidos donde se puede sacrificar la *consistencia* de los datos frente a la disponibilidad y a la tolerancia a particiones de red (Teorema CAP). Se centra en la disponibilidad y la escalabilidad lineal, es decir, se mejora el rendimiento en relación al número de nodos que se encuentran en la red [72]. Esta orientado a entornos donde no existe un nodo maestro, es decir, entornos totalmente distribuidos donde los nodos se comunican por una red *P2P*.

Este software se distribuye bajo una licencia *Apache 2*, es una licencia open-source que implica que se puede utilizar para cualquier propósito de forma gratuita. La licencia *Apache 2* es una licencia permisiva ya que la modificación del software distribuido con la misma no tiene por qué mantener esta licencia, exceptuando las partes que no hayan sido modificadas. [73]

A diferencia de los gestores de bases de datos relacionales, *Cassandra* no sigue las propiedades de transacciones *ACID*, pero en su lugar ofrece transacciones *atómicas*, *aisladas* y transacciones *persistentes* con *consistencia eventual* o modificable, es decir, el usuario puede decidir el nivel de exigencia que quiere fijar sobre la consistencia de una transacción. [7]

La herramienta de realización de consultas es *CQLSH*, una herramienta de línea de comandos para interactuar con *Cassandra* utilizando *CQL* (*Cassandra Query Language*) [20], un lenguaje para la realización de consultas y manipulación de los datos muy similar a *SQL*. [72]

Existen drivers (*ODBC* y *JDBC* (*Java*)) con los cuales podemos realizar el acceso a la base de datos utilizando distintos lenguajes de programación, tales como: *Java*, *Python*, *NodeJS* (*JavaScript*), *Dart*, *C++*... [21]

Para la administración de los nodos de un clúster, *Cassandra* provee la herramienta *Nodetool*. Esta herramienta brinda la capacidad de realizar tareas de administración sobre los nodos (añadir un nodo a un clúster, modificar la configuración de un nodo...) tanto como las herramientas necesarias para realizar una monitorización sobre el mismo. [72]

El soporte técnico de este sistema no lo realiza *Apache*, si no que recae sobre empresas de terceros. Por ejemplo, *DataStax*²⁷ ofrece servicios profesionales, servicio técnico y despliegues de este *SGBD*. [24]

2.7 MongoDB

MongoDB es un sistema de base de datos *NoSQL*, orientado a documentos y de código abierto. Los datos se almacenan en estructuras de tipo *BSON*²⁸ con un esquema dinámico, facilitando y agilizando la integración de los

²⁵ *Amazon's Dynamo* un conjunto de técnicas que juntas pueden formar un sistema de almacenamiento estructurado de alta disponibilidad o un depósito de datos distribuidos.[74]

²⁶ *Google Code* fue una plataforma que proveía control de versiones, seguimiento de incidencias para proyectos open-source. Actualmente se encuentra en modo de solo lectura desde Agosto de 2015. [75]

²⁷<https://www.datastax.com/>

²⁸Especificación similar a *JSON*

datos en determinadas aplicaciones.

Se puede obtener de forma gratuita bajo la licencia de código abierto *AGPL*²⁹. También ofrece una licencia comercial que incluye distintas características como integración con *SASL*, *LADP* o *Kerberos*.^[62]

Entre las principales características del gestor^[61], se puede destacar la capacidad de consultas *ad hoc*, con búsqueda por campos, consulta de rangos y expresiones regulares; Su capacidad de indexar cualquier campo en un documento³⁰ y crear índices secundarios, aumentando la eficiencia de las consultas^[37]; Soporta la replicación *primario-copia*, siendo cada grupo de los mismos un *replica set*³¹; La posibilidad de escalar horizontalmente, utilizando *Sharding*³², permitiendo añadir nuevos nodos al sistema mientras está en funcionamiento.^[40]; Mediante *GridFS*³³, *MongoDB* puede ser utilizado como un sistema de archivos con balanceo de carga y tolerante a fallos, aprovechando las características de *MongoDB*. Por último, permite realizar consultas con *JavaScript*^[38], enviándolas a la base de datos para que sean ejecutadas.

Mongo no garantizaba las propiedades *ACID* multidocumento hasta su versión 4.0 (entonces se cumplían para un solo documento), aunque posee las siguientes limitaciones^[63]:

- Bloquea la base a nivel de documento en las escrituras, impidiendo realizar operaciones de escritura concurrentes en el mismo documento.
- El gestor retorna tras escribir en al menos una réplica, lo que no basta para garantizar la durabilidad ni la verificabilidad.^[33]
- Si supera los 100 GB de datos, aparecen problemas de rendimiento.^[34]

El gestor posee drivers oficiales para diferentes lenguajes de programación: *C*, *C++*, *C#*, */.NET*, *Java*, *JavaScript*, *Node.js*, *Perl*, *PHP*, *Python*, *Ruby*, *Scala*, *Delphi* y *C++ Builder*.

Un apartado importante de *MongoDB* es *Mongo Shell*^[41], una interfaz interactiva en *JavaScript* mediante la cual se puede consultar y modificar datos e incluso realizar operaciones de administración. Puede ser utilizado en *Windows*, *macOS* o *Linux*, lo que convierte a *MongoDB* en un gestor multiplataforma.

MongoDB proporciona una serie de productos en la nube para poder almacenar y gestionar instancias del mismo, como lo son *MongoDB Stitch*, *MongoDB Atlas Data Lake* o *MongoDB Cloud Manager*, entre otros.^[32]

Además, el gestor posee una gran serie guías para cada una de sus funcionalidades en su página web^[36], así como un soporte técnico al que se contactar mediante correo o vía telefónica³⁴.

2.8 Apache HBase

HBase es sistema gestor de bases de datos *NoSQL*, con almacenamiento basado en columnas. Fue desarrollado por *Powerset*³⁵ para procesar grandes cantidades de datos para realizar las búsquedas con lenguaje natural. Actualmente es un proyecto de alto nivel de *Apache Software Foundation*.^[69]

Apache HBase provee la capa superior del sistema de ficheros de *Apache Hadoop*³⁶, está inspirado en *Google BigTable*³⁷ y el sistema de ficheros de *Google*.

Al estar programado con *Java*, *Apache HBase* es una solución multiplataforma, aunque desde la documentación del *SGBD* no recomiendan utilizar *Windows* en entornos de producción. Uno de los requisitos para utilizar este sistema es tener *JDK* instalado (recomiendan *JDK8*).^[19]

²⁹Licencia pública general de Affero de *GNU*

³⁰Los índices son almacenados en una estructura Árbol-B

³¹Un *replica set* es un grupo de procesos *mongod* que mantienen en mismo conjunto de datos.^[39]

³²Método para distribuir datos a través de múltiples máquinas

³³Especificación para almacenar y consultar archivos que superan el límite de tamaño de documentos *BSON* de 18 MB.^[35]

³⁴<https://www.mongodb.com/contact>

³⁵*Powerset* es una compañía norteamericana que desarrolló un motor de búsqueda con lenguaje natural.^[71]

³⁶*Apache Hadoop* es un framework de software que permite a las aplicaciones trabajar con miles de nodos y petabytes de datos.^[68]

³⁷*BigTable* es un sistema gestor de bases de datos creado por *Google* con las características de ser distribuido, de alta eficiencia y con licencia privativa.^[70]

Apache HBase es un *SGBD* orientado al procesamiento de grandes cantidades de datos, es decir, está orientado al análisis de datos (*OLAP*).

Este *SGBD* pertenece a *Apache Software Foundation*, una compañía sin ánimo de lucro que desde 1999 ha desarrollado software open-source. *HBase* es un proyecto de alto nivel de *Apache*. Su última versión estable se liberó en abril de 2006. [69]

HBase se distribuye bajo una licencia *Apache 2*, que, como se ha explicado anteriormente, permite la utilización del software para cualquier propósito de forma gratuita. ***La licencia Apache 2 es una licencia permisiva ya que la modificación del software distribuido con la misma no tiene por qué mantener esta licencia, exceptuando las partes que no hayan sido modificadas. Esto se puede quitar si eso.*** [73]

Sobre la escalabilidad, este *SGBD* está orientado a un entorno distribuido. Al igual que *Apache Cassandra*, está orientado a la escalabilidad lineal, es decir, el rendimiento del sistema mejora de manera lineal respecto al número de nodos del sistema. [16]

A diferencia de los *SGBD* relacionales, *HBase* provee transacciones *ACID* a nivel de fila. Para ello hace uso de mecanismos de bloqueo por filas (*row level lock*) y un sistema de control de la concurrencia multiversión [6][28]. Es un mecanismo que utilizan los *SGBD* relacionales para permitir la concurrencia, es decir, el sistema gestor almacena distintas versiones del objeto con el que se está trabajando mientras alguien esté trabajando con alguna de ellas. [76]

Para el acceso a este *SGDB*, desde la página web, nos proporcionan una serie de conectores. También, podemos realizar el acceso utilizando *SQL* existe *Apache Phoenix*³⁸, que proporciona una capa de acceso a *HBase* así como el acceso con drivers *JDBC* y e integración con herramientas de inteligencia de negocio [69].

Para la realización de tareas como las copias de seguridad, *Apache HBase* provee herramientas integradas para ello. Esta copia de seguridad se guardará automáticamente en la ubicación especificada, pudiendo ser el propio *HDFS* donde está desplegada la máquina, otro *HDFS* de otro centro de datos, o un servicio en cloud compatible con *HDFS*. [18]

3 SGBD

3.1 Oracle

pedro

3.2 MySQL

hayk

3.3 PostgreSQL

chus

3.4 DB2

tambo

3.5 Microsoft SQL Server

pedro

3.6 Microsoft Access

chus

³⁸*Apache Phoenix* es una base de datos de código abierto, masivamente paralela con soporte *OLTP* para *Hadoop* utilizando *HBase* como servicio de almacenamiento. [57]

3.7 Caché

hayk

3.8 VoltDB

chus

3.9 Cassandra

pedro

3.10 MongoDB

tambo

3.11 HBase

hayk

References

- [1] Consultas parametrizadas. https://www.adrformacion.com/knowledge/ofimatica/consultas_parametricas_en_access.html. Consultado el 29/04/2020.
- [2] Data concurrency and consistency. <https://docs.oracle.com/en/database/oracle/oracle-database/19/cncpt/data-concurrency-and-consistency.html#GUID-1D60EFCC-03F4-4A04-B099-1B4DE5D02C47>. Consultado el 26/04/2020.
- [3] Desventajas y ventajas voltdb. https://nanopdf.com/download/principales-sgbd_pdf. Consultado el 29/04/2020.
- [4] Velocidad de voltdb. <https://itblogpress.wordpress.com/2014/02/25/voltdb-bases-de-datos-45-veces-mas-rapidas-que-los-tradicionales-sistemas-mysql-y-oracle/>. Consultado el 29/04/2020.
- [5] Voltdb. <https://www.voltdb.com/why-voltdb>. Consultado el 29/04/2020.
- [6] G. Chanan. Apache hbase internals: Locking and multiversion concurrency control. https://blogs.apache.org/hbase/entry/apache_hbase_internals_locking_and, Enero 2013. Consultado el 29/04/2020.
- [7] DataStack. About transactions and concurrency control. https://docs.datastax.com/en/cassandra-oss/2.1/cassandra/dml/dml_about_transactions_c.html. Consultado el 27/04/2020.
- [8] DB2. Db2 - compatibilidad. <https://www.ibm.com/support/pages/system-requirements-ibm-db2-linux-unix-and-windows>. Consultado el 29/04/2020.
- [9] DB2. Db2 - escalabilidad. <https://www.eltiempo.com/archivo/documento/MAM-698231>. Consultado el 29/04/2020.
- [10] DB2. Db2 - express-c. https://www.ibm.com/support/knowledgecenter/SSEPGG_9.7.0/com.ibm.db2.luw.qb.server.doc/doc/r0054460.html. Consultado el 29/04/2020.
- [11] DB2. Db2 - herramientas. https://www.ibm.com/support/knowledgecenter/es/SSEPEK_10.0.0/intro/src/tpc/db2z_imtools.html. Consultado el 29/04/2020.
- [12] DB2. Db2 - introduction. <https://www.ibm.com/es-es/products/db2-database>. Consultado el 29/04/2020.
- [13] DB2. Db2 - jdbc. https://publib.boulder.ibm.com/tividd/td/ITIM/SC32-1147-03/es_ES/HTML/im451inst_ux_ws34.htm. Consultado el 29/04/2020.
- [14] DB2. Db2 - propiedades acid. <https://www.ibm.com/support/pages/commit-db2-luw-database>. Consultado el 29/04/2020.
- [15] Devcenter. Mvcc. <https://devcenter.heroku.com/articles/postgresql-concurrency>. Consultado el 28/04/2020.
- [16] A. S. Foundation. Apache hbase - features. <https://hbase.apache.org/index.html>. Consultado el 28/04/2020.
- [17] A. S. Foundation. The apache incubator. <https://incubator.apache.org/>. Consultado el 27/04/2020.
- [18] A. S. Foundation. Backup and restore. <https://hbase.apache.org/book.html#backuprestore>. Consultado el 29/04/2020.
- [19] A. S. Foundation. Basic prerequisites. <https://hbase.apache.org/book.html#basic.prerequisites>. Consultado el 27/04/2020.
- [20] A. S. Foundation. The cassandra query language (cql). <https://cassandra.apache.org/doc/latest/cql/index.html#cql>. Consultado el 27/04/2020.
- [21] A. S. Foundation. Client drivers. https://cassandra.apache.org/doc/latest/getting_started/drivers.html#client-drivers. Consultado el 27/04/2020.

- [22] A. S. Foundation. Downloading cassandra. <https://cassandra.apache.org/download/>. Consultado 27/04/2020.
- [23] A. S. Foundation. Installing cassandra. https://cassandra.apache.org/doc/latest/getting_started/installing.html#installing-cassandra. Consultado el 27/04/2020.
- [24] A. S. Foundation. Thirdpartysolution. <https://cwiki.apache.org/confluence/display/CASSANDRA2/ThirdPartySupport>, Julio 2019. Consultado el 27/04/2020.
- [25] Gartner. Magic quadrant for operational database management systems. <https://www.gartner.com/doc/reprints?id=1-1XT9MEFJ&ct=191125&st=sb>, Noviembre 2019. Consultado el 26/04/2020.
- [26] informaticaparatunegocio. Capacidad access. <https://www.informaticaparatunegocio.com/blog/ventajas-desventajas-las-bases-datos-access/>. Consultado el 29/04/2020.
- [27] S. Insights. Developer survey results 2019. <https://insights.stackoverflow.com/survey/2019#technology--databases>, 2019. Consultado del 26/04/2020.
- [28] Q. Landing. Multiversion concurrency control in hbase. <https://hellokangning.github.io/en/post/multiversion-concurrency-control-in-hbase/>, Diciembre 2017. Consultado el 29/04/2020.
- [29] Microsoft. Especificaciones access. <https://support.office.com/es-es/article/especificaciones-de-access-0cf3c66f-9cf2-4e32-9568-98c1025bb47c>. Consultado el 29/04/2020.
- [30] Microsoft. Microsoft access licencia. https://www.microsoft.com/es-es/p/access/cfq7ttc0k7q8?=&ef_id=CjwKCAjwqJ_1BRBZEiwAv73uwNScKHTjdgEy6LznNrbv_YoYDtv_cQXDIQNLLegF7gcGVZtUZCPmVRoCj58QAvD_BwE%3aG%3as&OCID=AID2000750_SEM_xUz8JKCM&MarinID=sxUz8JKCM%7c431595526780%7cms+access%7ce%7cc%7c%7c62233861395%7ckwd-26962656&lnkd=Google_0365SMB_App&gclid=CjwKCAjwqJ_1BRBZEiwAv73uwNScKHTjdgEy6LznNrbv_YoYDtv_cQXDIQNLLegF7gcGVZtUZCPmVRoCj58QAvD_BwE&activetab=pivot:overviewtab. Consultado el 28/04/2020.
- [31] Microsoft. Microsoft access requisitos. https://www.microsoft.com/es-es/p/access/cfq7ttc0k7q8?=&ef_id=CjwKCAjwqJ_1BRBZEiwAv73uwNScKHTjdgEy6LznNrbv_YoYDtv_cQXDIQNLLegF7gcGVZtUZCPmVRoCj58QAvD_BwE%3aG%3as&OCID=AID2000750_SEM_xUz8JKCM&MarinID=sxUz8JKCM%7c431595526780%7cms+access%7ce%7cc%7c%7c62233861395%7ckwd-26962656&lnkd=Google_0365SMB_App&gclid=CjwKCAjwqJ_1BRBZEiwAv73uwNScKHTjdgEy6LznNrbv_YoYDtv_cQXDIQNLLegF7gcGVZtUZCPmVRoCj58QAvD_BwE&activetab=pivot:regionofsystemrequirementstab. Consultado el 28/04/2020.
- [32] MongoDB. Mongoddb - cloud. <https://docs.mongodb.com/cloud/>. Consultado el 29/04/2020.
- [33] MongoDB. Mongoddb - durability problem. <https://hackingdistributed.com/2013/01/29/mongo-ft/>. Consultado el 29/04/2020.
- [34] MongoDB. Mongoddb - escalability problem. <https://jaxenter.com/mongodb-mocked-after-posting-100gb-scaling-checklist-106822.html>. Consultado el 29/04/2020.
- [35] MongoDB. Mongoddb - gridfs. <https://docs.mongodb.com/manual/core/gridfs/>. Consultado el 29/04/2020.
- [36] MongoDB. Mongoddb - guides. <https://docs.mongodb.com/guides/>. Consultado el 29/04/2020.
- [37] MongoDB. Mongoddb - indexes. <https://docs.mongodb.com/manual/indexes/>. Consultado el 29/04/2020.
- [38] MongoDB. Mongoddb - javascript. <https://docs.mongodb.com/manual/core/server-side-javascript/index.html>. Consultado el 29/04/2020.
- [39] MongoDB. Mongoddb - mongod processes. <https://docs.mongodb.com/manual/reference/program/mongod/#bin.mongod>. Consultado el 29/04/2020.
- [40] MongoDB. Mongoddb - sharding. <https://docs.mongodb.com/manual/sharding/>. Consultado el 29/04/2020.
- [41] MongoDB. Mongoddb - shell. <https://docs.mongodb.com/manual/mongo/>. Consultado el 29/04/2020.

- [42] Oracle. Backing up the database. <https://docs.oracle.com/en/database/oracle/oracle-database/19/bradv/backing-up-database.html#GUID-030BA0C5-832E-42A1-A63A-5E2D768BABB9>. Consultado el 26/04/2020.
- [43] Oracle. Data concurrency and consistency. https://docs.oracle.com/cd/B19306_01/server.102/b14220/consist.htm#i5702. Consultado el 26/04/2020.
- [44] Oracle. Oracle 19. <https://www.oracle.com/es/database/technologies/>. Consultado el 26/04/2020.
- [45] Oracle. Oracle database 19c install and upgrade. <https://docs.oracle.com/en/database/oracle/oracle-database/19/install-and-upgrade.html>. Consultado el 26/04/2020.
- [46] Oracle. Oracle real application clusters (rac). <https://www.oracle.com/es/database/technologies/rac.html>. Consultado el 26/04/2020.
- [47] Oracle. Oracle software licensing basics. <https://www.oracle.com/a/ocom/docs/corporate/oracle-software-licensing-basics.pdf>, Enero 2020. Consultado el 26/04/2020.
- [48] Oracle. Oracle technology global price list. <https://www.oracle.com/assets/technology-price-list-070617.pdf>, Marzo 2020. Consultado el 26/04/2020.
- [49] PGADMIN. Pgadmin. <https://www.pgadmin.org/>. Consultado el 28/04/2020.
- [50] PostgreSQL. Postgresql license. <https://www.postgresql.org/about/licence>. Consultado el 28/04/2020.
- [51] tecnologías información. Especificaciones access. <https://www.tecnologias-informacion.com/access.html>. Consultado el 29/04/2020.
- [52] TodoPostgres. Alta disponibilidad. <https://todopostgresql.com/alta-disponibilidad-postgresql/>. Consultado el 28/04/2020.
- [53] TodoPostgres. Hot-standby. <https://todopostgresql.com/tecnica-hot-standby>. Consultado el 28/04/2020.
- [54] TodoPostgres. Ventajas y desventajas. <https://todopostgresql.com/ventajas-y-desventajas-de-postgresql/>. Consultado el 28/04/2020.
- [55] WIKI. Microsoft access. https://es.wikipedia.org/wiki/Microsoft_Access. Consultado el 28/04/2020.
- [56] WIKI. Sloni-i. <https://en.wikipedia.org/wiki/Slony-I>. Consultado el 28/04/2020.
- [57] Wikipedia. Apache phoenix. https://en.wikipedia.org/wiki/Apache_Phoenix. Consultado el 29/04/2020.
- [58] Wikipedia. Arquitectura de nada compartido. https://en.wikipedia.org/wiki/Shared-nothing_architecture. Consultado el 29/04/2020.
- [59] Wikipedia. Db2 - purexml. <https://es.wikipedia.org/wiki/PureXML>. Consultado el 29/04/2020.
- [60] Wikipedia. Kubernetes. <https://es.wikipedia.org/wiki/Kubernetes>. Consultado el 29/04/2020.
- [61] Wikipedia. MongoDB - características principales. https://es.wikipedia.org/wiki/MongoDB#Caracter%C3%ADsticas_principales. Consultado el 29/04/2020.
- [62] Wikipedia. MongoDB - licencias y soportes. https://es.wikipedia.org/wiki/MongoDB#Concesi%C3%B3n_de_licencias_y_soporte. Consultado el 29/04/2020.
- [63] Wikipedia. MongoDB - principales limitaciones. https://es.wikipedia.org/wiki/MongoDB#Principales_limitaciones. Consultado el 29/04/2020.
- [64] Wikipedia. Motor jet. https://en.wikipedia.org/wiki/Microsoft_Jet_Database_Engine. Consultado el 29/04/2020.
- [65] Wikipedia. Newsq. <https://en.wikipedia.org/wiki/NewSQL>. Consultado el 29/04/2020.
- [66] Wikipedia. Postgresql. <https://es.wikipedia.org/wiki/PostgreSQL>. Consultado el 28/04/2020.

- [67] Wikipedia. VoltDB. <https://en.wikipedia.org/wiki/VoltDB>. Consultado el 29/04/2020.
- [68] Wikipedia. Apache hadoop. https://es.wikipedia.org/wiki/Apache_Hadoop, Agosto 2019. Consultado el 27/04/2020.
- [69] Wikipedia. Apache hbase. https://es.wikipedia.org/wiki/Apache_HBase, Octubre 2019. Consultado el 27/04/2020.
- [70] Wikipedia. Bigtable. <https://es.wikipedia.org/wiki/BigTable>, Julio 2019. Consultado el 27/04/2020.
- [71] Wikipedia. Powerset (company). [https://en.wikipedia.org/wiki/Powerset_\(company\)](https://en.wikipedia.org/wiki/Powerset_(company)), Diciembre 2019. Consultado el 27/04/2020.
- [72] Wikipedia. Apache cassandra. https://es.wikipedia.org/wiki/Apache_Cassandra, Abril 2020. Consultado el 27/04/2020.
- [73] Wikipedia. Apache license. https://es.wikipedia.org/wiki/Apache_License#Condiciones_de_la_licencia, Enero 2020. Consultado el 27/04/2020.
- [74] Wikipedia. Dynamo (storage system). [https://en.wikipedia.org/wiki/Dynamo_\(storage_system\)](https://en.wikipedia.org/wiki/Dynamo_(storage_system)), Febrero 2020. Consultado el 27/04/2020.
- [75] Wikipedia. Google developers - google code. https://en.wikipedia.org/wiki/Google_Developers#Google_Code, Abril 2020. Consultado el 27/04/2020.
- [76] Wikipedia. Multiversion concurrency control. https://es.wikipedia.org/wiki/Multiversion_concurrency_control#Ejemplo, Febrero 2020. Consultado el 29/04/2020.
- [77] Wikipedia. Oracle corporation. https://es.wikipedia.org/wiki/Oracle_Corporation, Febrero 2020. Consultado el 26/04/2020.
- [78] Wikipedia. Oracle database. https://en.wikipedia.org/wiki/Oracle_Database, Enero 2020. Consultado el 26/04/2020.