

# Análisis de Oracle, Cassandra y SQL Server

Pedro Tamargo

29 de abril de 2020

## Índice

Agradecimientos	1
1. Oracle	1
2. Apache Cassandra	3
3. Apache HBase	4
Referencias	5

## Agradecimientos

Muchas gracias a *Juan José Tambo Tambo*. Tus padres no son primos pero eres **MUY** tonto.

# 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* [1]. 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*” [2]. 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*<sup>1</sup>, 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* [3].

*Oracle* es una de las bases de datos líderes del mercado en bases de datos operacionales [4]. Cuenta con mucha documentación disponible desde su página web<sup>2</sup> y además cuenta con una comunidad muy activa en la resolución de incidencias y dudas en páginas como *StackOverflow* [5].

El dominio de aplicación de este *SGBD* se encuentra desde el desarrollo de aplicaciones hasta servicios de almacenes de datos [6]. **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*”<sup>3</sup> y ese es el número de “*Processor Licenses*” necesarios. Para el cálculo de “*Named User Licenses*” se aplicará el máximo entre:

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

Donde, *NumUserMinimum* se corresponde con el número de usuarios mínimos por cada “*Processor License*” [7]. Los precios de “*Processor License*” y “*Named User Licenses*” en la versión *Enterprise Edition* se corresponden con: 47500\$ y 950\$ [8]. 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. [9]

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.[10] Los mecanismos de bloqueo utilizados por este *SGBD* se dividen en dos grupos, bloqueos exclusivos y bloqueos compartidos. [11]

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*<sup>4</sup> y otras opciones de apoyo al desarrollo es posible configurando (si fuera necesario) el driver de conexión a la base de

---

<sup>1</sup>[https://es.wikipedia.org/wiki/Edgar\\_Frank\\_Codd](https://es.wikipedia.org/wiki/Edgar_Frank_Codd)

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

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

<sup>4</sup><https://dbeaver.io/>

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 [12].

El soporte técnico de este *SGBD* se adquiere por separado, y se relaciona con cada uno de los productos adquiridos. [8] **COMPROBAR ESTO**

## 2. 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*<sup>5</sup> 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*<sup>6</sup>. 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* [15]). En Febrero de 2010 se graduó como un proyecto de alto nivel. [16]

*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*. [17]

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. [18]

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 [16]. 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. [19]

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. [20]

La herramienta de realización de consultas es *CQLSH*, una herramienta de línea de comandos para interactuar con *Cassandra* utilizando *CQL* (*Cassandra Query Language*) [21], un lenguaje para la realización de consultas y manipulación de los datos muy similar a *SQL*. [16]

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++*... [22]

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. [16]

El soporte técnico de este sistema no lo realiza *Apache*, si no que recae sobre empresas de terceros. Por ejemplo, *DataStax*<sup>7</sup> ofrece servicios profesionales, servicio técnico y despliegues de este *SGBD*. [23]

---

<sup>5</sup>*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.[13]

<sup>6</sup>*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. [14]

<sup>7</sup><https://www.datastax.com/>

### 3. Apache HBase

*HBase* es sistema gestor de bases de datos *NoSQL*, con almacenamiento basado en columnas. Fue desarrollado por *Powerset*<sup>8</sup> para procesar grandes cantidades de datos para realizar las búsquedas con lenguaje natural. Actualmente es un proyecto de alto nivel de *Apache Software Foundation*. [25] *Apache HBase* provee la capa superior del sistema de ficheros de *Apache Hadoop*<sup>9</sup>, está inspirado en *Google BigTable*<sup>10</sup> 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*). [28]

*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. [25]

*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.*** [19]

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. [29]

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 [30][31]. 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. [32]

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*<sup>11</sup>, que proporciona una capa de acceso a *HBase* así como el acceso con drivers *JDBC* y e integración con herramientas de inteligencia de negocio [25].

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*. [34]

---

<sup>8</sup>*Powerset* es una compañía norteamericana que desarrolló un motor de búsqueda con lenguaje natural.[24]

<sup>9</sup>*Apache Hadoop* es un framework de software que permite a las aplicaciones trabajar con miles de nodos y petabytes de datos.[26]

<sup>10</sup>*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.[27]

<sup>11</sup>*Apache Phoenix* es una base de datos de código abierto, masivamente paralela con soporte *OLTP* para *Hadoop* utilizando *HBase* como servicio de almacenamiento. [33]

## Referencias

- [1] Wikipedia, "Oracle database." [https://en.wikipedia.org/wiki/Oracle\\_Database](https://en.wikipedia.org/wiki/Oracle_Database), Enero 2020. Consultado el 26/04/2020.
- [2] "Oracle corporation." [https://es.wikipedia.org/wiki/Oracle\\_Corporation](https://es.wikipedia.org/wiki/Oracle_Corporation), Febrero 2020. Consultado el 26/04/2020.
- [3] 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.
- [4] 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.
- [5] S. Insights, "Developer survey results 2019." <https://insights.stackoverflow.com/survey/2019#technology--databases>, 2019. Consultado del 26/04/2020.
- [6] Oracle, "Oracle 19." <https://www.oracle.com/es/database/technologies/>. Consultado el 26/04/2020.
- [7] 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.
- [8] Oracle, "Oracle technology global price list." <https://www.oracle.com/assets/technology-price-list-070617.pdf>, Marzo 2020. Consultado el 26/04/2020.
- [9] Oracle, "Oracle real application clusters (rac)." <https://www.oracle.com/es/database/technologies/rac.html>. Consultado el 26/04/2020.
- [10] Oracle, "Data concurrency and consistency." [https://docs.oracle.com/cd/B19306\\_01/server.102/b14220/consist.htm#i5702](https://docs.oracle.com/cd/B19306_01/server.102/b14220/consist.htm#i5702). Consultado el 26/04/2020.
- [11] "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.
- [12] 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.
- [13] 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.
- [14] Wikipedia, "Google developers - google code." [https://en.wikipedia.org/wiki/Google\\_Developers#Google\\_Code](https://en.wikipedia.org/wiki/Google_Developers#Google_Code), Abril 2020. Consultado el 27/04/2020.
- [15] A. S. Foundation, "The apache incubator." <https://incubator.apache.org/>. Consultado el 27/04/2020.
- [16] Wikipedia, "Apache cassandra." [https://es.wikipedia.org/wiki/Apache\\_Cassandra](https://es.wikipedia.org/wiki/Apache_Cassandra), Abril 2020. Consultado el 27/04/2020.
- [17] A. S. Foundation, "Installing cassandra." [https://cassandra.apache.org/doc/latest/getting\\_started/installing.html#installing-cassandra](https://cassandra.apache.org/doc/latest/getting_started/installing.html#installing-cassandra). Consultado el 27/04/2020.
- [18] A. S. Foundation, "Downloading cassandra." <https://cassandra.apache.org/download/>. Consultado 27/04/2020.
- [19] Wikipedia, "Apache license." [https://es.wikipedia.org/wiki/Apache\\_License#Condiciones\\_de\\_la\\_licencia](https://es.wikipedia.org/wiki/Apache_License#Condiciones_de_la_licencia), Enero 2020. Consultado el 27/04/2020.

- [20] DataStack, "About transactions and concurrency control." [https://docs.datastax.com/en/cassandra-oss/2.1/cassandra/dml/dml\\_about\\_transactions\\_c.html](https://docs.datastax.com/en/cassandra-oss/2.1/cassandra/dml/dml_about_transactions_c.html). Consultado el 27/04/2020.
- [21] A. S. Foundation, "The cassandra query language (cql)." <https://cassandra.apache.org/doc/latest/cql/index.html#cql>. Consultado el 27/04/2020.
- [22] A. S. Foundation, "Client drivers." [https://cassandra.apache.org/doc/latest/getting\\_started/drivers.html#client-drivers](https://cassandra.apache.org/doc/latest/getting_started/drivers.html#client-drivers). Consultado el 27/04/2020.
- [23] A. S. Foundation, "Thirdpartysolution." <https://cwiki.apache.org/confluence/display/CASSANDRA2/ThirdPartySupport>, Julio 2019. Consultado el 27/04/2020.
- [24] Wikipedia, "Powerset (company)." [https://en.wikipedia.org/wiki/Powerset\\_\(company\)](https://en.wikipedia.org/wiki/Powerset_(company)), Diciembre 2019. Consultado el 27/04/2020.
- [25] Wikipedia, "Apache hbase." [https://es.wikipedia.org/wiki/Apache\\_HBase](https://es.wikipedia.org/wiki/Apache_HBase), Octubre 2019. Consultado el 27/04/2020.
- [26] Wikipedia, "Apache hadoop." [https://es.wikipedia.org/wiki/Apache\\_Hadoop](https://es.wikipedia.org/wiki/Apache_Hadoop), Agosto 2019. Consultado el 27/04/2020.
- [27] Wikipedia, "Bigtable." <https://es.wikipedia.org/wiki/BigTable>, Julio 2019. Consultado el 27/04/2020.
- [28] A. S. Foundation, "Basic prerequisites." <https://hbase.apache.org/book.html#basic.prerequisites>. Consultado el 27/04/2020.
- [29] A. S. Foundation, "Apache hbase - features." <https://hbase.apache.org/index.html>. Consultado el 28/04/2020.
- [30] G. Chanan, "Apache hbase internals: Locking and multiversion concurrency control." [https://blogs.apache.org/hbase/entry/apache\\_hbase\\_internals\\_locking\\_and](https://blogs.apache.org/hbase/entry/apache_hbase_internals_locking_and), Enero 2013. Consultado el 29/04/2020.
- [31] 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.
- [32] Wikipedia, "Multiversion concurrency control." [https://es.wikipedia.org/wiki/Multiversion\\_concurrency\\_control#Ejemplo](https://es.wikipedia.org/wiki/Multiversion_concurrency_control#Ejemplo), Febrero 2020. Consultado el 29/04/2020.
- [33] Wikipedia, "Apache phoenix." [https://en.wikipedia.org/wiki/Apache\\_Phoenix](https://en.wikipedia.org/wiki/Apache_Phoenix). Consultado el 29/04/2020.
- [34] A. S. Foundation, "Backup and restore." <https://hbase.apache.org/book.html#backuprestore>. Consultado el 29/04/2020.