

Sistema de archivos, gestores de base de datos y Hadoop: ¿evolución o retroceso?

María del Pilar Ángeles

Resumen

La evolución de los sistemas de información ha estado marcada por cambios en el procesamiento, tiempo de respuesta, tipo y cantidad de información. En un principio, la necesidad era la de automatizar la operación diaria de un negocio. Posteriormente, se requirieron de análisis para tomar decisiones estratégicas. Actualmente, se necesita predecir eventos o comportamientos futuros a partir de grandes cantidades de datos variados, provenientes de redes sociales, videos o correos electrónicos. Las propuestas más recientes podrían hacernos cuestionarnos por qué estamos regresando al punto de partida, si en los sesenta iniciamos con los sistemas de archivos, que luego evolucionaron a los gestores de base de datos relacionales, los cuales solucionaron diversos problemas de inseguridad y falta de consistencia en los datos.

El presente artículo pone de manifiesto los cambios en las tecnologías de los sistemas de información conforme a las necesidades de las organizaciones, reflexiona sobre si éstos representan una evolución o un retroceso, y sugiere soluciones tecnológicas de acuerdo con cada necesidad, dado que no siempre lo que está a la moda es lo que se necesita.

Palabras clave: sistemas transaccionales y analíticos, inteligencia de negocios, datos masivos, ciencia de datos, bases de datos en memoria.

FILE SYSTEM, DATABASE MANAGERS AND HADOOP, EVOLUTION OR RETROGRADE?

Abstract

The evolution of information systems has been marked by changes in processing, response time, and type and quantity of information. In the beginning, the need was to automate the daily operation of a business. Subsequently, analysis was required to make strategic business decisions. Currently, it is necessary to predict future events or behaviors from large amounts of varied data from social media, videos, or emails. The most recent proposals could make us think why are we returning to the starting point, if we started with file systems and then evolved to relational database managers, which solved various problems of insecurity and lack of data consistency.

This article highlights the changes in information systems technologies according to the needs of organizations, reflects on whether these represent an evolution or a setback, and suggests technological solutions according to each given need, because that in style is not always what is required.

Keywords: transactional and analytical systems, business intelligence, big data, data science, in memory databases.

Recepción: 15/01/2020. Aprobación: 09/06/2021.

doi: <http://doi.org/10.22201/cuaieed.16076079e.2021.22.6.6>



María del Pilar Ángeles

pilarang@unam.mx
orcid.org/0000-0002-1754-2514
cienciadatos.iimas.unam.mx/profesores/pilarang/

Profesor Titular B de tiempo completo, Ingeniera en Computación, Maestra en Ciencias y *Ph.D. in Computer Science*. Sus áreas de interés y experiencia, por más de 25 años, son principalmente en calidad e integración de datos en ambientes federados, de-duplicación, proveniencia de datos, manejo de datos maestros, minería de datos y un poco de *big data* y ciencia de datos.

Introducción

Existen varios factores que pueden ayudar a decidir qué tecnologías son las más adecuadas a cada necesidad empresarial. En primer lugar, las organizaciones deben innovar para ser más competitivas, por ende, requieren gran capacidad de análisis en el menor tiempo posible. En segundo lugar, la generación de datos es cada vez más variada y masiva. Finalmente, la falta de actualización y el surgimiento constante de diversas tecnologías hace más difícil la toma de decisiones.

Aquí haremos una breve reseña de la evolución de las tecnologías de información a fin de identificar sus características más relevantes y ver si algunas tendencias corresponden a una evolución o retroceso. Esta reseña podría ayudar a decidir cuál tecnología será más adecuada con base a lo que ofrece, la infraestructura y recursos humanos que se tiene, así como las necesidades de innovación y análisis.

Desarrollo y evolución de los sistemas de información

El procesamiento de información viene de tiempos anteriores a la creación de la primera computadora. Durante los sesenta, el procesamiento de información se automatizó debido a que el costo de las computadoras bajó y algunas compañías privadas las pudieron adquirir. Las pequeñas y medianas empresas rentaban capacidad de procesamiento a centros de cómputo externos para correr, por ejemplo, su nómina, y dependían completamente de ellos. Los sistemas eran rígidos e inflexibles, cualquier cambio en la ubicación o diseño de los datos implicaba modificar los programas.

En los setenta, [Edgar F. Codd propuso el modelo relacional](#) (Codd, 1970) y publicó una serie de reglas para la evaluación de sistemas gestores de *base de datos relacionales*¹ (Codd, 1985). El éxito posterior de dichos gestores fue, por un lado, la independencia entre programas y datos y, por otro, la introducción del *lenguaje SQL*² (Reddy, 2017).

Durante los ochenta, los manejadores de base de datos relacionales se posicionaron en todos los sectores de la industria. Cada departamento desarrollaba sus propios sistemas transaccionales u *OLTP*³ para satisfacer sus necesidades. De esta manera, se crearon diversos silos de información dentro de la misma organización. Los manejadores de base de datos relacionales podían garantizar que la información permaneciera segura y consistente en todo momento a través de las *propiedades ACID*⁴ (HTML Rules, 2017).

Posteriormente, los directivos se dieron cuenta de la importancia que la información tenía para el negocio. Por ejemplo, si se deseaba saber las ventas de un cierto artículo en los últimos 10 años, requerían consolidar sus *silos de información* en una bodega de datos o *data warehouse*⁵ para poder tomar decisiones mediante un análisis de nivel empresarial (conocido como *sistemas OLAP*⁶ o analíticos) (Kimball, 1996; Inmon, 2002). Para ello había que extraer la información de los diversos sistemas, transformarla para adecuarla y cargarla (ETL)⁷ a una base de datos relacional como bodega de datos, y así poder analizar la información desde diferentes enfoques de negocio a lo largo del tiempo (LeapFrogBI, 2013).

El primer problema de los sistemas *OLAP* fue que el tiempo para el proceso ETL podía tardar varias horas y los reportes para los directivos no estaban listos a tiempo. Era común que llegara un director a preguntar por las ventas del día anterior y la respuesta fuera: —Sigue corriendo el proceso nocturno... El director volteaba por la ventana y decía: —¿Nocturno?, ¡si el sol ya salió! El segundo problema fue que, al almacenarse años de información para detectar tendencias y comportamientos, el disco y la memoria se empezaron a agotar.

Más recientemente surgieron los *sistemas manejadores de bases de datos columnares*⁸ (Moore, 2011), que permitían el proceso de ETL a la bodega de datos en una base de datos columnar. Con ello, el tiempo para resolver consultas *OLAP*

¹ Un gestor o manejador de base de datos relacionales es un software que se encarga de administrar y almacenar contiguos los registros o renglones de un conjunto de datos en disco.

² SQL o *Structured Query Language*, por sus siglas en inglés, es un lenguaje de programación muy fácil de aprender.

³ Los sistemas *OLTP* (*Online Transactional Processing*, por sus siglas en inglés) son aplicaciones de operación diaria que actualizan información, como las transacciones bancarias o las compras en un supermercado. Por lo general insertan, borran o actualizan datos.

⁴ ACID. Propiedades de atomicidad, consistencia, aislamiento y durabilidad, que garantizan que los datos siempre estén actualizados, consistentes y almacenados en disco. Estas propiedades son típicas de *OLTP*.

⁵ Base de datos o repositorio en donde se almacena información histórica para su análisis.

⁶ *On Line Analytical Processing* son aplicaciones que analizan la información para la toma de decisiones de negocio. Realizan principalmente lecturas, no actualizan datos.

⁷ *Extract, Transform, Load*. Proceso de extracción, transformación y carga a una bodega de datos para análisis *OLAP*.

⁸ Software que administra y almacena contiguos los valores de cada columna en una base de datos columnar.

se redujo considerablemente (Informática, 2020). Sin embargo, para las organizaciones ya no es suficiente conocer qué sucedió con sus ventas. Ahora se desea predecir o influenciar en las compras de los clientes; a esto se le llama *análisis predictivo y prescriptivo* (cuando te inducen a comprar cierta película porque te la sugieren). Lo anterior se logra con la introducción de otras disciplinas como la estadística computacional, el aprendizaje de máquina o la minería de datos (Han, Kamber y Pei, 2012; Be a better dev, 2020; ver figura 1).

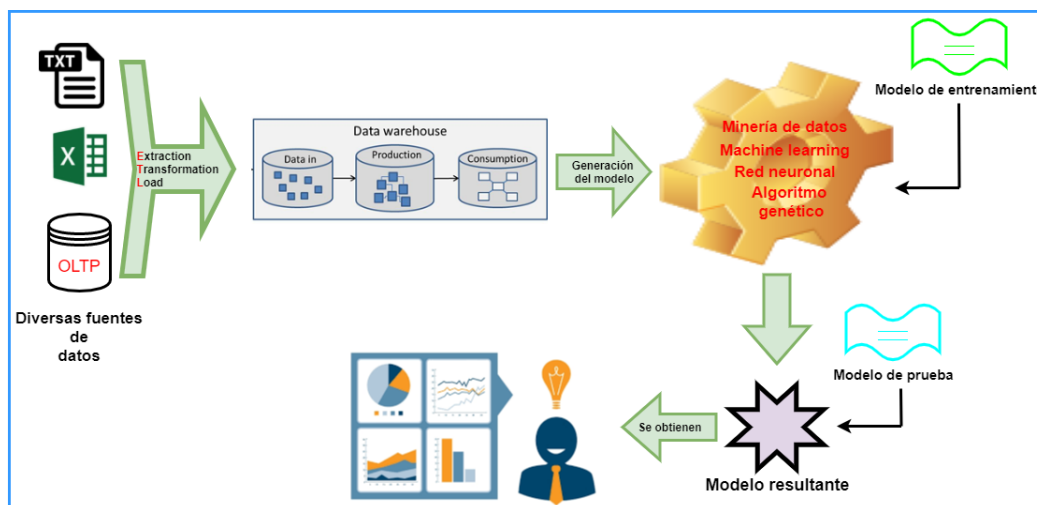


Figura 1. Arquitectura de inteligencia de negocios con análisis descriptivo, predictivo por gran variedad de herramientas para análisis.

El internet ha facilitado la interacción de las personas y ha contribuido al aumento de datos, tanto en variedad como cantidad, como tuits, publicaciones, videos, voz, datos geoespaciales, etcétera, los cuales hay que almacenar, administrar y analizar. Por tanto, las tecnologías han echado mano de técnicas como el *cómputo paralelo masivo* (MPP)⁹. Tal es el caso de [Teradata](#), [Oracle Real Application Cluster](#) y más recientemente [Hadoop](#). Estas innovaciones promueven el surgimiento de las *bases de datos multi-modelo*¹⁰ y la *tecnología NoSQL* (Sadalage y Fowler, 2013). Esta última permite, por ejemplo, almacenar más fácilmente la información de redes sociales distribuida en varias computadoras (Simply explained, 2020). La desventaja es que los datos permanecen inconsistentes por ciertos períodos de tiempo (Panicker, 2016). Lo cual, nos hace reflexionar una vez más, acerca de lo que realmente necesitamos: ¿rapidez en el manejo de datos o precisión y consistencia en ellos?

Hoy en día, las organizaciones necesitan analizar diversos tipos de información, lo más rápido posible, a nivel predictivo y prescriptivo. Por ejemplo, ya no es viable tener que esperar horas a que se realice el ETL de correos electrónicos y tuits a registros en una bodega de datos para su posterior análisis. Esto da lugar al problema del análisis de *datos masivos*¹¹ (Katsov, 2013). Como posibles soluciones al análisis de datos masivos se han propuesto diversas tecnologías, como las bases de datos multimodelo o el marco de trabajo Hadoop¹² (Defog Tech, 2019; Hillam, 2012; Borthakur, 2010).

⁹Tipo de arquitectura que usa muchas CPUs separadas corriendo en paralelo para procesar un sólo programa.

¹⁰Manejador de base de datos que puede almacenar la información de diversas formas, como columnar, relacional, llave-valor, etcétera.

¹¹Conocido como *big data* en inglés, consiste en el manejo de grandes cantidades de información de diversos tipos, como correos electrónicos, imágenes, hojas de cálculo, etcétera.

¹²Marco de trabajo que usa un sistema de archivos para almacenar y procesar simultáneamente datos distribuidos, usando los recursos de disco, procesamiento y memoria del grupo de computadoras donde residen los datos.

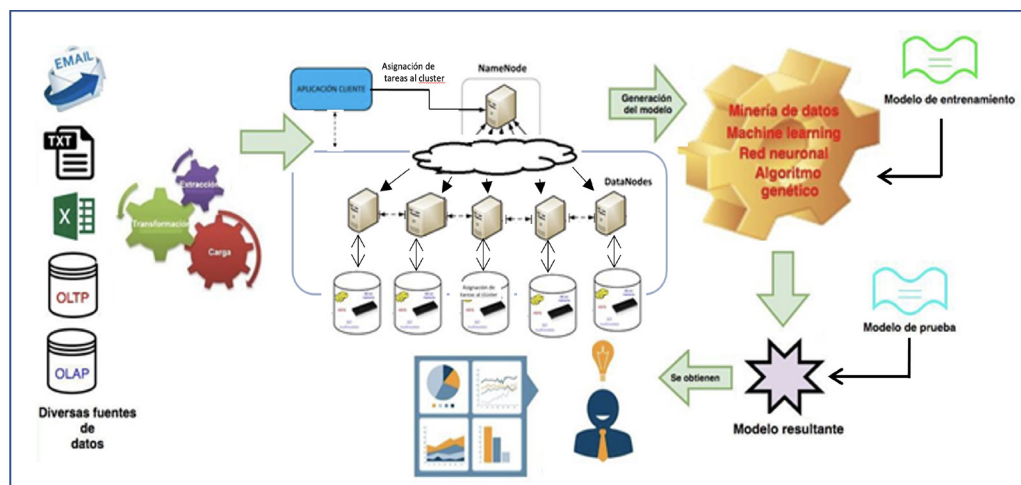


Figura 2. Arquitectura para análisis descriptivo y/o predictivo de cualquier tipo de dato con bases de datos en memoria, multimodelo o sistema de archivos distribuido Hadoop.

En el siglo XXI

El [Instituto de Investigación Hasso Plattner](#) y la [compañía sap A. G.](#) anunciaron en 2012 un *software* que maneja base de datos en *memoria principal*:¹³ [SAP HANA DB](#) (Plattner, 2014), y que soportaba datos geoespaciales, grafos y texto dentro del mismo sistema de almacenamiento. HANA DB puede correr sistemas transaccionales, analíticos y reduce el tiempo de procesamiento (Knapp, 2018).

Durante la primera década del siglo *xxi*, la tendencia fue incorporar *softwares* de análisis estadístico e inteligencia artificial al manejador de base de datos (conocido como *in-database analytics*), donde la programación reside en la propia base de datos (Looker, 2017). Aquí sería pertinente hacer notar el regreso a las desventajas de los años sesenta al juntar datos y programas.

En 2018, la empresa Intel anunció sus módulos de memoria persistente [Optane dc](#) (Alcorn, 2018). Su característica principal es la capacidad de no perder datos al cortarse el flujo de la energía eléctrica. Esto aumenta la confianza en el uso de las bases de datos en memoria, por la seguridad y rapidez. Actualmente estas tecnologías se ofrecen en la nube, haciéndolas más asequibles para todo mundo.

En la actualidad, la tendencia en los sistemas manejadores de bases de datos es administrar cualquier tipo de información y mejorar el tiempo de respuesta en aplicaciones OLAP y OLTP (ver figura 3). No obstante, el regresar al almacenamiento en sistemas de archivos distribuidos en varios discos y que sean leídos simultáneamente (en paralelo) con Hadoop (Borthakur, 2010), o Amazon S3, entre otros, no parece ser una gran ventaja. Sobre todo, si se considera que no podrá ser un sistema en tiempo real, no es fácil de implementar y pierde rendimiento con el uso excesivo de red y disco (Dilan, 2013).

¹³ Componente de una computadora que consta de circuitos integrados que trabajan a altas velocidades para el almacenamiento temporal de datos. Se conoce también como memoria interna y consiste en ROM y RAM.

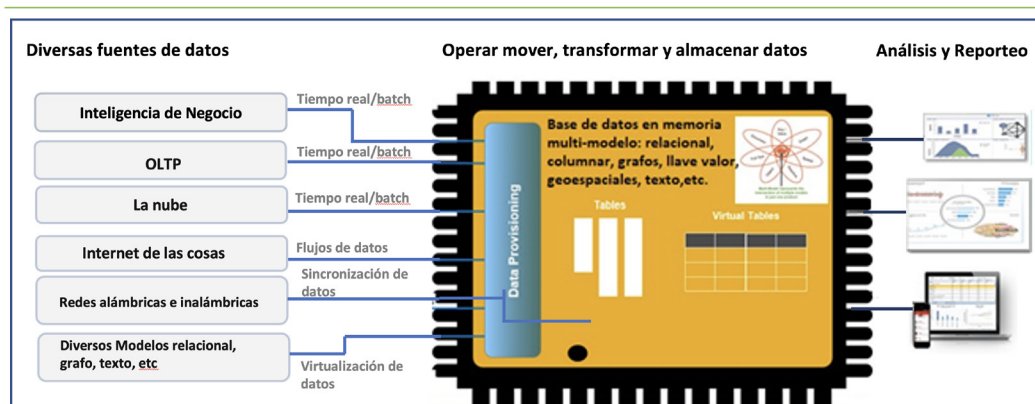


Figura 3. Gestor de bases de datos multimodelo en memoria persistente para soporte de procesamiento por lotes, en línea o en tiempo real para operación o análisis de información.

Otra alternativa para analizar datos rápidamente es procesarlos conforme se van generando (Psaltis, 2017). Esto se conoce como *análisis de ráfagas* de datos o *streaming analytics* en inglés. Por ejemplo, podemos generar una alerta al recibir datos provenientes de un sensor, si detectamos que se ha excedido algún límite en la medición. Estos sistemas consultan poca información simultáneamente, su objetivo principal es, por ejemplo, leer datos de un sensor, y tomar decisiones y acciones rápidamente. Hay que considerar que el análisis de ráfagas de datos sería complicado y no recomendable cuando se necesita adecuar los datos antes del análisis y éste es complejo.

En cuanto a la complejidad en el análisis, la *ciencia de datos* ha sido el resultado de la evolución de diversas tecnologías que explican, descubren o predicen fenómenos a partir de cualquier tipo de dato. Sin embargo, para ello se requiere de personal altamente capacitado y de una infraestructura robusta si es que se manejan grandes cantidades de información (Great Learning, 2019).

Si no se cuenta o no se puede mantener una infraestructura robusta ni personal capacitado, el cómputo en la *nube*¹⁴ parecería la panacea, pues se puede contratar un amplio portafolio de soluciones de *hardware* y *software* para dar respuesta a la demanda, sin tener que preocuparse por los detalles técnicos. Sin embargo, toda la infraestructura en la que se guardan los datos ya no reside en las propias oficinas y probablemente tampoco en las del proveedor de la nube, sino en un tercero, y esto puede constituir en un obstáculo de proporciones gigantescas en el plan de continuidad de negocio de la empresa.

El pago de los servicios en la nube se puede establecer con base en los recursos y funciones que se necesitan para operar. Si un mes no se paga la factura a tiempo, la posibilidad de que la compañía se quede sin acceso a sus propios datos y aplicaciones es real, similar a lo que pasaba en la década de los sesenta con el pago del tiempo compartido. Aquí nos preguntamos: ¿evolución o retroceso? Además, hay que considerar que cuanto mayor sea la empresa, la cantidad de personas y número de proyectos, más grande será la factura y más cerca se podría estar de un problema que podría llevar a la empresa a situaciones muy difíciles ante la imposibilidad de acceder a sus datos (Hodges, 2019).

¹⁴ Paradigma que permite ofrecer servicios de computación como almacenamiento, *software* y capacidad de cómputo a través de una red, sin una gestión activa directa por parte del usuario.

Reflexión

Dentro de las principales propuestas para el análisis rápido de grandes cantidades de cualquier tipo de información (datos masivos o *big data*) están el marco de trabajo de Hadoop y la tecnología NoSQL¹⁵ (Sadalage y Fowler, 2013), en los que no se garantiza al 100 por ciento la seguridad y consistencia que ya se tenían en tecnologías anteriores.

Si a inicios de los sesenta se tenían sistemas de archivos y a lo largo de décadas se evolucionó a modelos relacionales, modelos columnares y en memoria, que siguen soportando seguridad, persistencia y consistencia, ¿cómo es que las propuestas actuales las garantizan en tiempo real? ¿Esto implica evolución o retroceso?

En primer lugar, las tecnologías NoSQL y Hadoop son generalmente de código abierto y soportan cualquier tipo de información como videos, textos o imágenes que dan mayor riqueza al análisis. Dada la naturaleza del procesamiento de este tipo de información, las operaciones típicas son de lectura, así que tampoco es imprescindible que soporten consistencia e integridad. Además, están surgiendo tecnologías NoSQL multimodelo con dicho soporte, para el desarrollo de sistemas de información transaccionales, que puedan almacenar y manejar diversos formatos (Be a better dev, 2020).

En segundo lugar, los sistemas de bases de datos multimodelo en memoria son más robustos, requieren licencia y son tecnologías dirigidas al sector industria, con suficiente presupuesto para migrar varios *terabytes*¹⁶ o *petabytes*¹⁷ de información. El problema es el costo de memoria principal. Sin embargo, recordemos que están surgiendo RAM persistentes y que se espera abaratar este recurso computacional. Así que esta tecnología también podrá soportar grandes cantidades de ráfagas de datos, sistemas analíticos y transaccionales.

Entonces, el retroceso consiste en no estar conscientes de las ventajas y desventajas que cada tecnología ofrece y empezar un proyecto de operación diaria, consolidación o análisis de información sin considerar las capacidades, infraestructuras y características de volumen, variedad, velocidad y tipo de análisis. Así como utilizar una tecnología específica sólo porque es el tema principal en las redes sociales o está de moda (Lumen, 2016).

Recomendaciones

Si alguna tecnología está de moda, no implica que la compañía deba implementarla para verse innovadora, pues lo más importante es que se vea qué necesidades de operación o análisis se tienen.

Si se necesitan describir las tendencias claves en los datos existentes y esto se puede lograr sólo consultando las fuentes de datos existentes, un sistema de tipo OLAP nos permitirá un análisis básico descriptivo como proporciones, tasas, razones o promedios.

¹⁵ Bases de datos que administran información en grafos, columnas, pares llave-valor, documentos o cualquier tipo de dato semiestructurado.

¹⁶ Un terabyte son 1.024 Gigabytes o 1.099.511.627.776 bytes.

¹⁷ Un petabyte son 1.024 terabytes o más de un cuatrillón de bytes.

Si dependemos de un análisis estadístico para obtener información nueva o histórica y el utilizar ésta para predecir patrones de comportamiento y aplicarlos a eventos del pasado, presente o futuro, lo que se requiere es un sistema de información predictivo. El cómo implementarlo dependerá de la rapidez deseada para el análisis y la toma de decisiones, así como los tipos y cantidades de información.

Si la información no contiene textos, grafos, imágenes, audio o video, por ejemplo, entonces, con una solución de inteligencia de negocio que contemple una buena herramienta de análisis a través de estadística y aprendizaje de máquina sería suficiente.

Si la cantidad de datos crecerá rápidamente, es posible que el rendimiento y, por ende, el proceso de análisis y toma de decisiones empiece a degradarse con el tiempo. En consecuencia, es conveniente usar una base de datos columnar y en memoria, o bajo una arquitectura paralela distribuida, sin tener que llegar al uso de un sistema de archivos.

Si la mayoría de los datos son, por ejemplo, video, texto, grafos, imágenes y no serán cantidades masivas de información (como terabytes), lo más conveniente es implementar una solución de inteligencia de negocio que almacene los datos en tecnología NoSQL para ahorrar tiempo en transformar y “no tener que estructurar” los datos.

Si la mayoría de los datos son video, imágenes, textos, se prevé un aumento masivo en el volumen de los datos y no se requiere una respuesta en tiempo real, puede usarse una *arquitectura clúster de computadoras*¹⁸ con alguna base de datos NoSQL.

Si la mayoría de los datos provienen de sensores calibrados y por ende no requieren adecuarse previamente para la predicción se puede utilizar análisis de ráfagas de datos para una toma de decisiones prácticamente en tiempo real. El almacenamiento en este caso puede ser opcional.

Para concluir, existen tantas tecnologías y tendencias en el mercado de los sistemas de información que podríamos sentirnos abrumados y no saber qué tipo de proyecto será el mejor. Lo importante es saber cuál es la estrategia de negocio, con qué tipo de información y recursos contamos, qué tipo de análisis y tiempo de respuesta deseamos y decidir acorde a ello.

Referencias

- ❖ Alcorn, P. (2018). Intel Displays 512GB Optane DC Persistent Memory DIMMs. *Tom's Hardware*. <https://cutt.ly/fRO6rkD>
- ❖ Be a Better Dev. (2020, 10 de febrero). *sql vs Nosql Explained* [video]. YouTube. <https://www.youtube.com/watch?v=ruz-vK8lesE>

¹⁸ Los *clusters* son grupos de computadoras que forman una red de alta velocidad y que se comportan como si fuesen un único servidor. Tienen a cada nodo realizando la misma tarea, controlada y planificada por *software*. Es la arquitectura sobre la que trabaja Hadoop y NoSQL para el proceso paralelo y distribuido.

- ❖ Borthakur, D. (2007). The Hadoop Distributed File System: Architecture and Design The Apache Software Foundation.
 - ❖ Codd, E. (1985). Is Your DBMS Really Relational? *ComputerWorld*, 9(41).
 - ❖ Codd, E. (1970). A Relational Model of Data for Large Shared Data Banks. *Communications of the ACM*.
 - ❖ Defog Tech. (2019, 17 de abril). *Google File System - Paper that inspired Hadoop* [video]. YouTube. <https://www.youtube.com/watch?v=eRgFNW4QFDcv>
 - ❖ Dilan, G. (2013, 13 de septiembre).. What are the disadvantages of mapreduce? *Stack overflow*. <https://cutt.ly/CRPwTQX>
 - ❖ Great Learning. (2019, 18 de octubre). *What is Data Science?* [video]. YouTube. https://www.youtube.com/watch?v=Nrfht_c3T7w
 - ❖ Han, J., Kamber, M. y Pei, H. (2012). *Data mining, Concepts and techniques*. Morgan Kaufmann.
 - ❖ Harmon, A. L. (1959, 13 de mayo). [Carta a P. Z. Ingerman]. Archive of Computer History Museum. <https://cutt.ly/WRPwFQi>
 - ❖ HDFS Architecture. (2021, 25 de octubre). <https://cutt.ly/qRPwNa>
 - ❖ Hillam, J. (2012, 14 de julio). *Intricity 101. What is Hadoop?* [video]. YouTube. <https://www.youtube.com/watch?v=9s-vSeWej1U>
 - ❖ Hodges, R. (2019, 30 de julio). Far more than cloud: Thoughts on the future of database management systems. *Altinity*. <https://cutt.ly/ORPw73p>
 - ❖ HTML Rules. (2017, 11 de octubre). *Transacciones y el test ACID en bases de datos* [video]. YouTube. <https://www.youtube.com/watch?v=SaUai23Z3Tc>
 - ❖ *Informática para tu negocio*. (2020). Fundamentos de una base de datos columnar. <https://cutt.ly/mRPedAa>
 - ❖ Inmon, W. (2002). *Building the Data Warehouse*. John Wiley & Sons.
 - ❖ *Intel latam*. (2020). Memoria persistente Intel® Optane™. <https://cutt.ly/9RPegVb>
 - ❖ *Intel latam-optane*. (2020). Intel® Optane™ DC Persistent Memory Partner: Accenture. <https://cutt.ly/ERPen5Y>
 - ❖ Katsov, I. (2013, 20 de agosto). In-Stream Big Data Processing. *Highly Scalable Blog*. <https://highlyscalable.wordpress.com/2013/08/20/in-stream-big-data-processing/>
 - ❖ Kimball, R. (1996). *The Data Warehouse Toolkit: Practical Techniques for Building Dimensional Data Warehouses*. John Wiley & Sons.
 - ❖ Knapp, B. (2018, 30 de octubre). *IBM Cloud. What is SAP HANA?* [video]. YouTube. <https://www.youtube.com/watch?v=8VXurKENGRE>
 - ❖ LeapFrogBI. (2013, 18 de marzo). *Dimensional Modeling: Introduction* [video]. YouTube. <https://www.youtube.com/watch?v=cwpl-3rkRYQ>
-

- ❖ Looker. (2017, 18 de diciembre). *Analytical Databases: Differentiators of Database Technologies* [video]. Youtube. <https://www.youtube.com/watch?v=XDjcvUKiNqc>
- ❖ Lumen. (2016, 14 de enero). *Data Lakes: Hadoop Vs. In-Memory Databases*. <https://blog.lumen.com/data-lakes-hadoop-vs-in-memory-databases/>
- ❖ Moore, T. (2011). *The Sybase IQ Survival Guide*. Lulu.com.
- ❖ Panicker, M. (2016, 17 de junio). *Big Data and NoSQL Databases Tutorial (BaSE)- Part 3* [video]. YouTube. <https://www.youtube.com/watch?v=pRgUGkDxj7k>
- ❖ Parmar, Y. (2017, 7 de enero). *Relational databases, Advantages over flat file systems* [video]. YouTube. <https://www.youtube.com/watch?v=vKkdhOj2Oog>
- ❖ Plattner, H. (2014). *A Course in- Memory Data Management*. Springer-Verlag.
- ❖ Psaltis, A. (2017). *Streaming Data: Understanding the real-time pipeline*. Manning Publications Co.
- ❖ Reddy, G. (2017, 18 de enero). *Database vs file system* [video]. YouTube. <https://www.youtube.com/watch?v=y3dc6Bjq2LM>
- ❖ Sadalage, P. y Fowler, M. (2013). *NoSQL Distilled: A Brief Guide to the Emerging World of Polyglot Persistence*. Pearson Education, Inc.
- ❖ SAP Inside Track. (2019, 2 de septiembre). *Databases go multimodel* [video]. YouTube. <https://www.youtube.com/watch?v=LE0BXZV2m6s>
- ❖ Simply explained. (2020, 8 de diciembre). *How do NoSQL databases work? Simply Explained!* [video]. YouTube. <https://www.youtube.com/watch?v=0buKQHokLK8>
- ❖ Vikramtakka. (2013, 18 de marzo). *3 - ETL Tutorial | Extract Transform and Load* [video]. YouTube. <https://www.youtube.com/watch?v=WZw0OTgCBOY>

CÓMO CITAR ESTE ARTÍCULO

- ❖ Ángeles, María del Pilar. (2021, noviembre-diciembre). Sistema de archivos, gestores de base de datos y Hadoop: ¿evolución o retroceso? *Revista Digital Universitaria (RDU)*, 22(6). <http://doi.org/10.22201/cuaieed.16076079e.2021.22.6.6>