

Impacto positivo de las bases de datos documentales en las Medianas Empresas

Carlos M. Fonseca Gómez & Nicolas R. Miranda Ladino

Resumen— Las bases de datos no relacionales surgen como alternativa de las bases de datos tradicionales mejorando la escalabilidad y con un mayor performance y disponibilidad, es por esto por lo que el objetivo del artículo es demostrar el impacto positivo respecto al análisis de datos al tener una base de datos no relacional, para lo cual se compararan los tiempos de respuesta para las 4 funciones básicas en bases de datos relacionales y NoSQL utilizando 39224 registros obtenidos de Twitter. Con esto demostraremos que este sistema de almacenamiento de información puede ser una mejor opción al momento de tener grandes volúmenes de información.

Palabras Claves— Análisis de datos, bases de datos no relacionales, grandes volúmenes de datos, NoSQL.

I. INTRODUCCIÓN

DEBIDO al crecimiento y complejidad de los datos en las aplicaciones de la actualidad, tales como Facebook, Google, Amazon, entre otras, surge la necesidad de un nuevo sistema en el cual se puedan alojar y consultar cantidades masivas de datos de manera más rápida y eficiente, este sistema se denomina bases de datos NoSQL[1], el cual deja a un lado características importantes de las bases de datos tradicionales entre ellas la atomicidad, aislamiento y durabilidad y se reemplazan por rendimiento, escalabilidad y concurrencia [2], donde gran cantidad de usuarios interactúan con información de la misma, las cuales no utilizan una estructura de datos en forma de tablas y relaciones sino que se usan otro tipo de formatos como clave – valor, documentales, mapeo de columnas o grafos.

Este tipo de sistemas no relacionales tiene grandes ventajas en primer lugar el aspecto de rendimiento debido a que se ejecutan en máquinas con pocos recursos para su adecuado funcionamiento y si se desea escalar su operación al trabajar de manera horizontal se añaden nodos y se muestra cuales están disponibles para sentencias de ejecución como de consulta, lo cual este tipo de características convierte este sistema de bases de datos en una opción muy accesible para las empresas en especial de tamaño mediano por su costo e inversión y porque este tipo de empresas tienen un crecimiento notorio en unos de sus principales activos el cual es la información por lo tanto es necesidad vital el procesamiento, alojamiento y análisis de grandes volúmenes de datos.

II. MATERIALES Y MÉTODOS

Hace algunos años se utilizaban las bases de datos relacionales en la mayoría de los proyectos que exigen la persistencia de los datos, no obstante, los esquemas relacionales pueden no ser los más apropiados cuando se tiene un alto volumen de datos y se le da prioridad a la disponibilidad [3]. Es por esto por lo que este artículo tiene como finalidad demostrar el impacto positivo que tienen las bases de datos NoSQL al momento de que se disponga de grandes volúmenes de información.

Para comprobar esto se utilizaron datos obtenidos de Twitter ya que este nos da la posibilidad de trabajar con sus datos utilizando la documentación encontrada en la página <http://dev.twitter.com>. Se obtuvo la información gracias a un programa realizado en Python el cual generó una base de datos con 39224 registros en formato JSON tal como se puede evidenciar en la revista de tecnologías de la información [4].”

III. DESARROLLO

Las bases de datos SQL deben cumplir el concepto de ACID (automáticas, consistentes, aisladas y durables), lo que significa que el modelo relacional debe garantizar que una transacción se ejecute completamente o no se ejecute (automáticas), los datos están en estado válido o inválido (consistentes), cada transacción es independiente entre sí (aisladas) y al terminar una transacción la información perdura en el tiempo (durables) [5]. Es por esto por lo que negocios que deban garantizar su disponibilidad y requieren que las bases de datos sean replicables fácilmente y puedan hacer frente a los fallos en los centros de datos y a los picos de consumo en los servidores generen estrategias para gestionar esta gran cantidad de información tales como aumentar las características de procesamiento y de almacenamiento, lo cual resulta muy costoso y es por esto por lo que surgen nuevas formas de manejar la información como NoSQL.

NoSQL a diferencia de las bases de datos relacionales almacenan los datos sin mecanismos estructurados lo que facilita la gestión tanto de datos estructurados como los no estructurados. Algunas características de estas bases de datos son, la escalabilidad horizontal, la habilidad de distribución, el uso eficiente de los recursos, la libertad de esquema, el uso de modelos de concurrencia débil y las consultas simples [6]. La principal ventaja de NoSQL en comparación con las bases de datos relacionales es la de poder escalar el rendimiento de las

C. Mario Fonseca. Universidad Libre, Bogotá – Colombia
carlosm.fonseca@unilibrebog.edu.co

Nicolas R. Miranda. Universidad Libre, Bogotá – Colombia, nicolasr-mirandal@unilibre.edu.co

Corresponding author: C. Mario Fonseca

operaciones a lo largo de varios servidores interconectados para garantizar el rendimiento optimo de las operaciones.

Uno de los primeros presentadores de las bases de datos NoSQL para la gestión de grandes cantidades de información fue Google el cual presento Google Big Tables como solución para satisfacer tanto las necesidades de su explorador web como de Google Earth y otros servicios de la compañía [7].

Los modelos con los que cuentan las bases de datos NoSQL se clasifican en 4 grupos los cuales son:

- Modelo llave-valor
- Orientado a documentos
- Orientado a grafos
- Orientado en columnas

Modelo llave-valor

Estas son las más simples dentro de las bases de datos NoSQL en las cuales se le asigna una clave también llamada llave única a un valor el cual es una cadena de caracteres. En estas bases de datos la acción de encontrar el valor que este asociado a una clave se llama lookup y la relación entre una clave y su valor se llama correlación [8]. Estas bases de datos aseguran una buena escalabilidad y brindan la posibilidad de almacenar los datos en un ambiente distribuido, se recomienda utilizarlas en casos donde lo primordial sea la velocidad de consulta o si se poseen muchos datos que requieran ser procesados repetidas veces y tienen valores cambiantes.

Orientado a documentos

Estas están diseñadas para la gestión de información orientada a documentos o datos semiestructurados, estos sistemas están diseñados en base al concepto abstracto de documento [9]. Un documento es la unidad de almacenamiento de estas bases de datos, en las cuales toda la información almacenada se guarda en formato de documento. Este modelo permite manejar millones de lecturas simultaneas gracias a que posee una lectura simple ya que toda la información relacionada a una entidad se almacena en un solo documento.

Orientado a grafos

Estas bases de datos utilizan una estructura de grafos la cual tiene nodos los que serían objetos o entidades, aristas que indicarían la relación entre las entidades y las propiedades para representar y almacenar datos [10]. Este modelo brinda un sistema de almacenamiento el cual brinda libre indexado por adyacencia lo que significa que cada elemento contiene un puntero el cual direcciona a su elemento adyacente esto permite que no se requiera una búsqueda por índices. Es recomendado utilizarlas en caso de que se tengan estructuras dinámicas complejas.

Orientado en columnas

Las bases de datos orientadas a columnas almacenan los datos en columnas en lugar de filas como el modelo relacional, estos sistemas proporcionan un mejor rendimiento cuando se presentan casos en los que se necesite realizar una transformación de datos la cual incluya una gran parte de las filas, pero solo una o una pequeña parte de las columnas [11]. Estas bases de datos son recomendadas a empresas que posean datos cuya lectura sea mas frecuente que su escritura y en casos en los que se realicen demasiadas operaciones con los atributos de las entidades.

Las bases de datos NoSQL se adaptan perfectamente a muchas aplicaciones modernas ya que estas ofrecen esquemas flexibles los cuales permiten un desarrollo más rápido e interactivo haciendo que estas sean ideales para datos semiestructurados y no estructurados [12], estas bases de datos están diseñadas para escalar usando clústeres distribuidos de hardware en lugar de escalar añadiendo servidores caros y solidos además estas están optimizadas para modelos de datos específicos permitiendo un mayor rendimiento que las bases de datos relacionales.

A pesar de todo estas bases de datos no son siempre la mejor opción pues cuando lo que se busca es la consistencia de los datos sin dar posibilidad al error lo mejor seria utilizar una base de datos SQL [13], pero las bases de datos NoSQL son una solución óptima cuando los datos que tenemos son variables, si lo que buscamos es capturar y procesar eventos, cuando tenemos gran cantidad de datos de modo lectura y si no se tiene el presupuesto para grandes máquinas y se debe recurrir a unas de menor rendimiento.

A continuación, podemos ver una grafica tomada de la revista de tecnologías de la información [4], en la cual se puede evidenciar la diferencia que hay en los tiempos de respuesta entre una base de datos relacional como MySQL y una NoSQL como MONGODB.

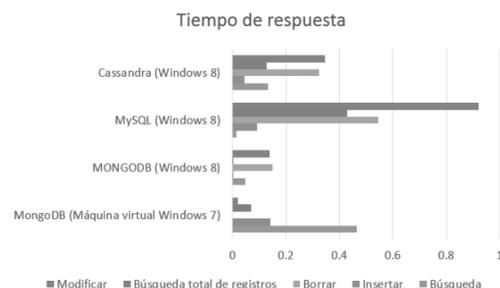


Figura 1. Tiempos de respuesta

Como podemos ver en la figura 1 los tiempos de respuesta son mejores en las bases de datos NoSQL cuando se realizan búsquedas de gran cantidad de registros, mientras que las

relacionales son mejores al momento de realizar la búsqueda de un solo registro.

Se realizó la comparación de precios entre los servicios que ofrece MONGODB y AWS [14], como se puede ver en las figuras 2, 3, 4, 5 y 6 los servicios prestados por las bases de datos NoSQL como MONGODB son más económicos.

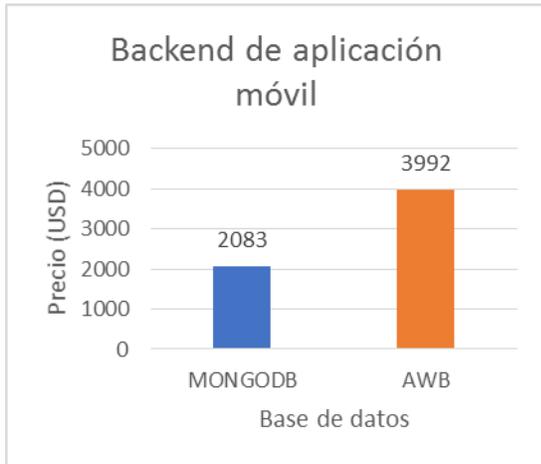


Figura 2: Backend de aplicación móvil

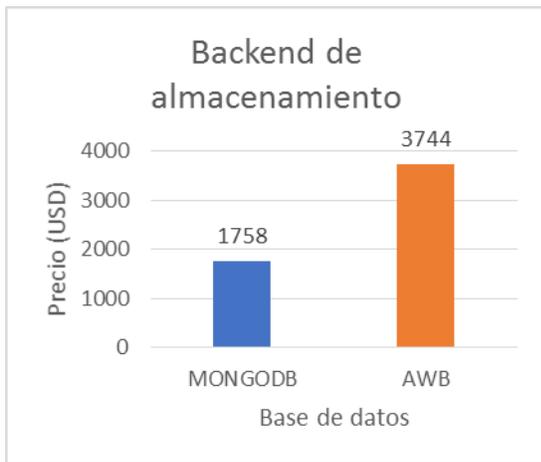


Figura 3: Backend de almacenamiento

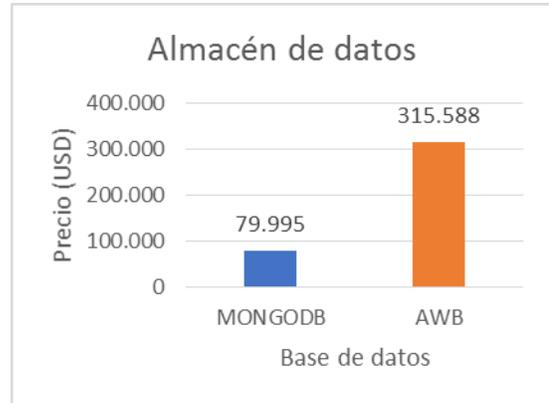


Figura 4: Almacén de datos

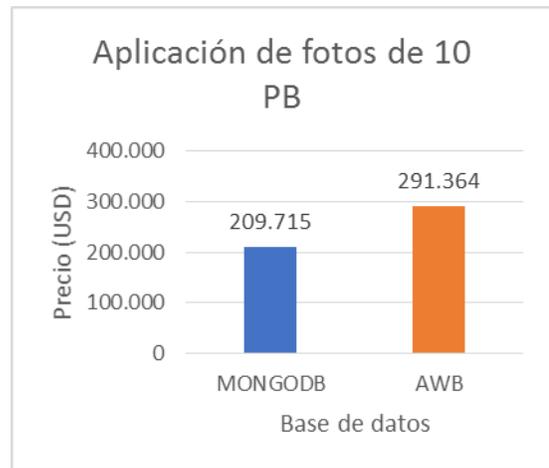


Figura 5: Aplicación de fotos de 10 PB

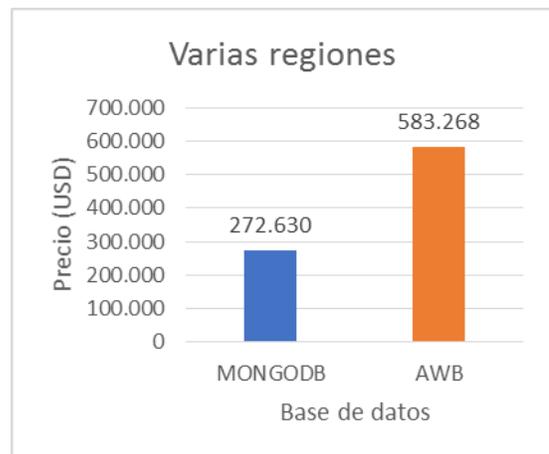


Figura 6: Varias regiones

IV CONCLUSIONES

A pesar de que las bases de datos relacionales están presentes en la mayor parte de las organizaciones no son las únicas en aportar una solución a los nuevos retos a los que se están enfrentando las empresas.

NoSQL no soluciona todos los problemas de las bases de datos relacionales existen varios modelos y cada uno de estos soluciona algunos problemas específicos es por esto que las empresas deben conocer muy bien su negocio y sus necesidades para saber que estructura utilizar al momento de almacenar sus datos.

Las bases de datos NoSQL ofrece diversas ventajas y con un bajo costo de infraestructura lo que permite a las empresas medianas más oportunidades de crecimiento y evolución.

Referencias

- [1] Pollo Cattaneo, M., López Nocera, M., & Daián Rottoli, G. (1). Rendimiento de tecnologías NoSQL sobre cantidades masivas de datos. CUADERNO ACTIVA, (6), 11-17. Recuperado a partir de <http://ojs.tdea.edu.co/index.php/cuadernoactiva/article/view/197>
- [2] (Strauch et al., 2011). <http://ojs.tdea.edu.co/index.php/cuadernoactiva/article/view/197/199>
- [3] Herrera, H., & Rueda Valenzuela, C. (2016). NoSQL, la nueva tendencia en el manejo de datos. Tecnología Investigación Y Academia, 4(1), 147-150. Recuperado a partir de <https://revistas.udistrital.edu.co/ojs/index.php/tia/article/view/8649>
- [4] Montes, F., González, D. & Campoy, A. (2016). Experiencia en el manejo de tecnología NoSQL contrastada con el modelo relacional usando software libre. http://www.ecorfan.org/bolivia/researchjournals/Tecnologias_de_la_Informacion/vol3num9/Revista_de_tecnologias_de_la_informacion_V3_N9_2.pdf
- [5] Colorado Pérez, M. (2017). NoSQL, ¿es necesario ahora?. Tecnología Investigación Y Academia, 5(2), 174-179. Recuperado a partir de <https://revistas.udistrital.edu.co/ojs/index.php/tia/article/view/8771>
- [6] Ramírez Arévalo, H. H., & Herrera Cubides, J. F. (2013). UN VIAJE A TRAVÉS DE BASES DE DATOS ESPACIALES NoSQL. Redes De Ingeniería, 4(2), 57-69. <https://doi.org/10.14483/2248762X.5923>
- [7] De la Hoz Freyle, J. (2015). Big data y NoSQL: su rol en la revolución del cloud computing y sus retos hacia la estandarización. I+D Revista de Investigaciones. Recuperado a partir de <http://www.udi.edu.co/revistainvestigaciones/index.php/ID/article/view/46>
- [8] Romero, A. C., Sanabria, J. S. G., & Cuervo, M. C. (2012). Utilidad y funcionamiento de las bases de datos NoSQL. Facultad de Ingeniería, 21(33), 21-32. <https://www.redalyc.org/pdf/4139/413940772003.pdf>
- [9] del Busto, H. G., & Enríquez, O. Y. (2013). Bases de datos NoSQL. Revista Telemática, 11(3), 21-33. <http://www.revistatelematica.cujae.edu.cu/index.php/tele/article/view/74/74>
- [10] Antiñanco, M. J. (2014). Bases de Datos NoSQL: Escalabilidad y alta disponibilidad a través de patrones de diseño (Doctoral dissertation, Facultad de Informática). <http://sedici.unlp.edu.ar/handle/10915/36338>
- [11] Herranz Gómez, R. (2014). Bases de datos NoSQL: arquitectura y ejemplos de aplicación (Master's thesis). <https://e-archivo.uc3m.es/handle/10016/22895#preview>
- [12] amazon. ¿Que es NoSQL? <https://aws.amazon.com/es/nosql/>
- [13] Javier. (2015). NoSQL vs SQL; principales diferencias y cuando elegir cada una de ellas. <https://blog.pandorafms.org/es/nosql-vs-sql-diferencias-y-cuando-elegir-cada-una/>
- [14] Google. (2016). Precios <https://cloud.google.com/pricing/?hl=es>