

## **Análisis de aplicabilidad de bases de datos tipo GRAFOS en comparación a las tradicionales SQL**

Madrid Julio Edgar  
CC 1143135727  
Código estudiantil: 201811290144  
Correo: [edgar.madrid@unisimonbolivar.edu.co](mailto:edgar.madrid@unisimonbolivar.edu.co)

Sulbaran Soto Valentina  
CC: 1001779400  
Código estudiantil: 201811292820  
Correo: [valentina.sulbaran@unisimonbolivar.edu.co](mailto:valentina.sulbaran@unisimonbolivar.edu.co)

Vargas Contreras Edwin  
CC: 72177917  
Código estudiantil: 201811289627  
Correo: [edwin.vargas@unisimonbolivar.edu.co](mailto:edwin.vargas@unisimonbolivar.edu.co)

Zabala Roqueme Carlos  
CC 1143460015,  
Código estudiantil: 201711282074  
[carlos.zabala@unisimonbolivar.edu.co](mailto:carlos.zabala@unisimonbolivar.edu.co)

Trabajo de Investigación del Programa INGENIERÍA DE SISTEMAS

Tutor:  
**Diana Judith Heredia Vizcaino**

**Antecedentes:**

Uno de los principales problemas de las organizaciones es poder disponer de herramientas que le permitan almacenar, gestionar gran cantidad de información, situación que ya se está quedando compleja con las bases de datos relacionales, por lo que este artículo se centró en estudiar las funcionalidades de las bases de datos tipo grafo que, mejoren el rendimiento de este manejo.

### **Objetivos:**

- Introducir al uso de una base de datos tipo grafos no relacional, en busca de encontrar fluidez al momento de realizar consultas.
- Poder manejar grandes cantidades de datos de manera fluida, sin generar retrasos en los procesos de consulta.
- Analizar el rendimiento de un lector de base de datos tipo grafos.

### **Materiales y Métodos:**

Para el uso del aplicativo, empleamos de un dispositivo que cuente con un sistema operativo en aptas condiciones para soportar, cargues de información, consultas de datos y modificaciones en el aplicativo. Se utilizan conocimientos básicos en lenguaje cypher, para realizar manejo de datos en la base de datos.

### **Resultados:**

Para evaluar las funcionalidades de la base de datos NEO4J, debemos comentar que, el proceso de instalación fue supremamente fácil, no hubo inconvenientes, ya en sí la parte del uso de la herramienta para realizar las consultas no es relativamente fácil, puesto que el modo de las consultas se aleja un poco de lo tradicional a lo que normalmente las personas que utilizan bases de datos relacionales están acostumbrados, de este modo se tiene que recurrir a manuales e instructivos que ayuden a hacer el cambio de chips, representando esto un tiempo de aprendizaje más largo para los aprendices.

En la parte de muestra de resultados de las consultas realizadas, debemos anotar que a simple vista no es muy funcional la parte de grafos, porque no es visible todos los grafos que se puedan obtener en una consulta cuando es una cantidad muy grande de registros retornados, ya se tendría que analizar este despliegue con otra herramientas como Python que, se conecte a esta base de datos y poder ver otra forma de mostrar los datos, pero esta prueba no estaba en el alcance de la investigación.

Ya para la parte de la prueba piloto de la propuesta de investigación, desarrollada sobre el versión desktop y las configuraciones del PCs que no son propiamente de un servidor robusto, teniendo en cuenta estas características, se procede a describir la primera métrica evaluada que, fué la carga de los archivos, los cuales tenían cantidades de filas distintas en cada uno, por lo que se prefirió analizar el resultado solamente de los archivos “consulta” y “pacientes”, que representaban los archivos con la mayor cantidad de filas, siendo así, que para el caso del archivo de “consulta”, con 101310 filas tuvo una duración en la carga de 77 minutos, y para el caso del archivo de pacientes con 99343 filas tomó un tiempo de 44 minutos, considerándose un tiempo demasiado lenta y que a medida que se realizaban nuevas cargas el rendimiento se veía más retrasado que el anterior.

Continuando con el análisis de las métricas, se diseñaron una serie de consultas que, nos proporcionaron información sobre rendimiento y versatilidad de las funciones que tiene disponible la base de datos en su entorno para la extracción de datos almacenados en los nodos, y sus relaciones.

### **Conclusiones:**

Si bien, estos resultados no son fáciles de leer en algunos casos, no se descarta la recomendación de utilizarla, si tenemos en cuenta que, las condiciones para las pruebas realizadas no eran las mejores, y que esas mismas lecturas realizadas con otros lenguajes de programación especializados en bigdata y analítica de datos, las lecturas y representaciones son más manejables y garantiza una experiencia de usuario más acomodada a la necesidad de cada una de las organizaciones que lo requieren.

**Palabras clave:** bigdata, base de datos, NOSQL, grafos, NEO4j.

## ABSTRACT

### **Background:**

### **Objective:**

### **Materials and Methods:**

### **Results:**

To evaluate the functionalities of the NEO4J database, we must comment that the installation process was extremely easy, there were no inconveniences, and in itself the part of using the tool to perform the queries is not relatively easy, since the mode of the queries is a bit far from the traditional to what people who use relational databases are used to, in this way they have to resort to manuals and instructions that help to change the chips, representing this a time of longer learning for learners.

In the sample part of the results of the queries made, we must note that at first glance the graph part is not very functional, because not all the graphs that can be obtained in a query are visible when there is a very large number of returned records. This deployment would already have to be analyzed with other tools such as Python that connects to this database and be able to see another way of displaying the data, but this test was not in the scope of the investigation.

For the part of the pilot test of the research proposal, developed on the desktop version and the configurations of the PCs that are not properly a robust server, taking into account these characteristics, we proceed to describe the first metric evaluated, which was the loading of the files, which had amounts of different rows in each one, so it was preferred to analyze the result only of the files "consultation" and "patients", which represented the files with the largest number of rows, being thus, which in the case of the "query" file, with 101,310 rows, had a loading duration of 77 minutes, and in the case of the patient file with 99,343 rows, it took a time of 44 minutes, considering a time too slow and that as new uploads were performed, performance was more lagged than before.

Continuing with the analysis of the metrics, a series of queries were designed that provided us with information on the performance and versatility of the functions that the database has available in its environment for the extraction of data stored in the nodes, and their relationships.

### **Conclusions:**

Although these results are not easy to read in some cases, the recommendation to use it is not ruled out, if we take into account that the conditions for the tests carried out were not the best, and that the same readings carried out with other programming languages specialized in big data and data analytics, the readings and representations are more manageable and guarantee a user experience more suited to the needs of each of the organizations that require it.

**KeyWords:** big data, database, NOSQL, graphs, NEO4j

## REFERENCIAS

- [1] Martínez Cedillo, M. (2019, Agosto). Base de datos orientada a Grafos¿Qué son y para qué se usan? Singular.[online].Disponible:<https://www.singular.com/es/que-son-bases-datos-a-grafos/>

- [2] Henderson, R. (2020, Julio). Using graph databases to detect financial fraud. 2020(7), 6 - 10.[online].Disponible:<https://ezproxy.unisimon.edu.co:2055/science/article/pii/S1361372320300737>
- [3] Bellini, P., & Nesi, P. (2018, Abril). Performance assessment of RDF graph databases for smart city services. Journal of visual languages & computing, 45(2018 The Authots), 24-38.[online].Disponible:<https://ezproxy.unisimon.edu.co:2055/science/article/pii/S1045926X1730246X>
- [4] Menor, R. (2020, Febrero). Que es RDF ? Que es RDF?.[online].Disponible:<https://www.seofreelance.es/que-es-rdf-introduccion-a-rdf/>
- [5] Batrouni, M., Bertaux, A., & Nicole, C. (2018, Mayo). Journals & Books. Scenario Analysis, from BigData to Black Swan, 28(2018 Elsevier Inc), 131-139.[online].Disponible:  
<https://ezproxy.unisimon.edu.co:2055/science/article/pii/S157401371730151X>
- [6] Mondal, S., Basu, A., & Mukherjee, N. (2020, Octubre). Building a trust-based doctor recommendation system on top of multilayer graph database. 110(2), 15.[online].Disponible:  
<https://ezproxy.unisimon.edu.co:2055/science/article/pii/S1532046420301775>
- [7] Camacho, E. (2021, Enero). NoSQL la evolución de las bases de datos. SG.[online].Disponible:  
<https://sg.com.mx/revista/28/nosql-evolucion-bases-datos>
- [8] Diaz, W. (2013, Mayo). Bases de datos NOSQL. BASES DE DATOS NOSQL: LLEGARON PARA QUEDARSE.[online].Disponible: <http://basededatosnosql.blogspot.com/2013/05/bases-de-datos-nosql-llegaron-para.html>
- [9] Ionos. (2019, Octubre). Graph database: bases de datos para una interconexión eficiente. Ionos.[online].Disponible:  
<https://www.ionos.es/digitalguide/hosting/cuestiones-tecnicas/graph-database/>
- [10] AWS. (2021, Febrero). ¿Qué es Amazon Neptune? AWS.[online].Disponible:[https://docs.aws.amazon.com/es\\_es/neptune/latest/userguide/intro.html](https://docs.aws.amazon.com/es_es/neptune/latest/userguide/intro.html)
- [11] Galiana, P. (2020, Enero). ¿Qué es SAP y para qué sirve? IEBS.[online].Disponible:<https://www.iebschool.com/blog/que-es-para-que-sirve-sap-management/>
- [12] Objectivity. (n.d.). InfiniteGraph. THINGSPAN.[online].Disponible:<https://objectivity.com/products/thingspan/>
- [13] Sparsity-technologies. (2015). Sparsity Technologies - Powering Extreme Data.[online].Disponible:  
<http://www.sparsity-technologies.com/#>.
- [14] Castañeda, J. (2016, Febrero). IDENTIFICACIÓN DE FACTORES TECNOLÓGICOS PARA LA APLICACIÓN DE BIG DATA. IDENTIFICACIÓN DE FACTORES TECNOLÓGICOS PARA LA APLICACIÓN DE BIG DATA. [online].Disponible:  
<http://repositorio.utp.edu.co/dspace/bitstream/handle/11059/6638/0057C346.pdf?sequence=1&isAllowed=y>