

Aspectos de Ingeniería de Software, Bases de Datos Relacionales, y Bases de Datos NoSQL para el desarrollo de Sistemas de Software Híbridos

Luciano Marrero , Pablo Thomas , Ariel Pasini , Rodolfo Bertone , Eduardo Ibáñez , Verena Olsowy , Fernando Tesone , Patricia Pesado 

Instituto de Investigación en Informática LIDI (III-LIDI)
Facultad de Informática – Universidad Nacional de La Plata
50 y 120 La Plata Buenos Aires
Centro Asociado CIC

{lmarrero, pthomas, apasini, pbertone, eibanez, vaguirre, volsowy, ftesone, ppesado}@lidi.info.unlp.edu.ar,

RESUMEN

Se presenta una línea de investigación que tiene por objeto estudiar las problemáticas que afronta la Ingeniería de Software y las Bases de Datos (Relacionales y No Relacionales) ante el avance en la comunicación digital, el acceso constante a la información (internet) y el aumento en la utilización de tecnología móvil, lo que genera que las aplicaciones sean de uso común por millones de usuarios simultáneamente.

El modelo relacional de bases de datos (Codd 1970), es, sin duda, el modelo predominante de almacenamiento de información. Sin embargo, la idea de considerar que un único modelo de datos pueda adaptarse de forma eficiente a todos los requerimientos ha sido discutida, y ha dado lugar a un conjunto de alternativas que plantean almacenar la información de forma no estructurada. Surgen así, nuevos Sistemas de Gestión de Bases de Datos que poseen sus propias implementaciones, no relacionales, y se

denominan Bases de Datos NoSQL (No solo SQL) [1, 2]

Este contexto requiere nuevas formas de pensar los diseños de Bases de Datos en complemento con las tradicionales metodologías ágiles para el desarrollo de Software.

Los principios o reglas que se aplican a un esquema de datos relacional no resultan apropiados para una Base de Datos No Relacional (NoSQL). Esto se debe a que NoSQL no representa un tipo de Base de Datos, sino que son un conjunto de tipos de Bases de Datos y cada una de ellos posee su propia forma de estructurar y almacenar sus datos [6, 8, 30].

El avance de la tecnología impulsó aspectos que hace algunos años no eran considerados en el desarrollo del Software, tales como, movilidad, geolocalización y la diversidad de los dispositivos electrónicos involucrados [5].

Palabras claves: Diseño no estructurado de datos, Metodologías ágiles, Procesos de diseño

de Bases de Datos Relacionales y Bases de Datos NoSQL, Bases de Datos como servicio en la nube, Aplicaciones Móviles.

CONTEXTO

Esta línea de Investigación forma parte del Proyecto *“Diseño, desarrollo y evaluación de sistemas en escenarios híbridos para áreas clave de la sociedad actual: educación, ciudades inteligentes y gobernanza digital”* del Instituto de Investigación en Informática LIDI de la Facultad de Informática, acreditado por el Ministerio de Educación de la Nación. Hay cooperación con Universidades de Argentina y se está trabajando con Universidades de Europa en proyectos financiados por el Ministerio de Ciencia y Tecnología de España y la AECID. Se utilizan los recursos de Hardware y Software disponibles en el III-LIDI para diseñar, desarrollar y probar diferentes soluciones a problemáticas relacionadas con escenarios a investigar. Como resultado de esto, se espera obtener métricas reales que sirvan como referencia para los investigadores en la comparación de resultados. Las publicaciones científicas generadas y la transferencia continua de resultados concretos, validan esta línea de investigación.

1. INTRODUCCIÓN

“Diseño, desarrollo y evaluación de sistemas en escenarios híbridos para áreas clave de la sociedad actual: educación, ciudades inteligentes y gobernanza digital” del Programa de Incentivos, es un proyecto que propone profundizar las investigaciones que se vienen realizando en el III-LIDI y extender la mirada a

nuevos desafíos y cambios que están en gestación. Se organiza en tres subproyectos que permiten atender de manera ordenada el objetivo general propuesto. *“SP1: Ingeniería de Software para escenarios híbridos”*, *“SP2: Ciudades Inteligentes sostenibles (CIS). Gobernanza Digital. Buenas prácticas y calidad”* y *“SP3: Creación de tecnologías digitales para el escenario educativo”*. [22].

Este artículo se centra en el subproyecto *“SP1 - Ingeniería de Software para escenarios híbridos”*. Se orienta a la investigación de metodologías y técnicas de la Ingeniería de Software, Bases de Datos (Relacionales y No Relacionales), estudio de procesos de diseño de Base de Datos No Relacionales, aplicaciones multiplataforma, escenarios híbridos y aplicaciones móviles, entre otras temáticas [22].

En la actualidad existen un conjunto de nuevos desafíos que van desde las etapas iniciales de diseño y continúan hasta que se obtiene el producto final..

En el caso de las Bases de Datos, cuando se trabaja con Bases de Datos no relacionales, generalmente se plantea directamente su estructura a nivel físico teniendo en cuenta el problema a resolver y la flexibilidad de los esquemas no estructurados de datos. Realizar un esquema para documentar la semántica de los datos en una Base de Datos NoSQL al igual de cómo se diseña una Base de Datos Relacional es algo que aún es tema de discusión.

Una especificación de requerimientos posee la ambigüedad propia del lenguaje natural, lo que introduce un cierto grado de subjetividad por parte del diseñador al momento de realizar el esquema de datos. Las reglas que se aplican en un modelo de datos relacional son, en

general, rígidas y estructuradas y que no consideran algunos aspectos que generalmente están presentes en un modelo de datos no estructurado perteneciente a una Base de Datos no relacional (NoSQL). En general, NoSQL prioriza la eficiencia y disponibilidad para un subconjunto de consultas que impactarán sobre el sistema [1, 7, 8].

Es importante remarcar que el diseño de una Base de Datos es un proceso cuyo objetivo es definir la estructura adecuada para el almacenamiento de datos de un sistema de información. El diseño influye de forma directa en la eficiencia para almacenar, recuperar y analizar los datos en la Base de Datos.

Para las Bases de Datos relacionales, se cuenta con un proceso de diseño adoptado el cual inicia con un modelo conceptual, luego se deriva a un modelo lógico relacional y finaliza con un modelo físico de datos.

Para el caso de las Bases de Datos NoSQL, no se cuenta un proceso masivamente aceptado para el diseño de esquemas de datos, sino que han surgido como soluciones directamente a nivel físico. No obstante, a medida que han adquirido popularidad, se han propuesto diferentes procesos y/o metodologías de diseño NoSQL, necesarios para comprender la semántica de los datos [28, 31].

Actualmente, muchas empresas desarrollan aplicaciones que son mundialmente utilizadas poniendo el foco en la escalabilidad de sus recursos y enfrentando nuevos retos para satisfacer la demanda de sus usuarios [2, 14].

Existe un amplio abanico de tecnologías, tanto en Bases de Datos (Relacionales y No Relacionales), como así también, en lenguajes para el desarrollo de todo tipo de aplicaciones de Software [10, 11, 12].

En el contexto de Base de Datos, existen distintos motores de Bases de Datos Relacionales y No Relacionales para el desarrollo de aplicaciones de propósito general, entre ellos, podemos encontrar a SQLite [16], Couchbase [17], Firebase Realtime Database [23], Google Cloud Firestore [23], Redis [24], MongoDB [25], Cassandra [26], Neo4j [27], entre otras.

Los motores de Bases de Datos son parte de la continua evolución tecnológica, esto genera que el usuario tenga un gran abanico de posibilidades para elegir el motor de Base de Datos que se adapte a sus exigencias [1, 9].

Otra temática de investigación se obtiene de analizar el proceso de despliegue del Software. Las PyMES en Argentina representan casi el 80% de la industria del software y dada la necesidad de ser competitivas deben mejorar sus métodos y procesos de trabajo. En la mayoría de las empresas el proceso de despliegue no se realiza de manera sistemática y controlada, esto impacta en la finalización del proyecto y la no aceptación del producto final, ocasionando inconvenientes que generan rehacer el trabajo y baja de productividad en su proceso. Ante estas dificultades, se propone realizar un modelo de proceso de despliegue de sistemas de software que permita a las PyMES mejorar la ejecución del proceso de despliegue [4, 18]. En este contexto, la comunicación y la sincronización entre los equipos de trabajo continúan siendo un pilar fundamental para el éxito de un proyecto. La utilización de repositorios de información y sincronización permiten realizar un control de versiones distribuido, trabajando en modo offline o en modo online, con la facilidad de disponer

herramientas específicas para la resolución de conflictos entre versiones [3, 4].

LÍNEAS DE INVESTIGACIÓN Y DESARROLLO

3. Estudio y análisis de procesos de diseño para Bases de Datos no relacionales [1, 2].
4. Aplicación y adaptación de procesos de diseño de Bases de Datos no relacionales en casos de usos particulares [28, 33].
5. Metodologías y técnicas de la Ingeniería de Software y su aplicación en el desarrollo de Software para escenarios híbridos [2, 30].
6. Desarrollo de casos de estudio y métricas de performance para distintos tipos de motores de Bases de Datos No Relacionales (NoSQL) orientados a Familia de Columnas y orientados a Grafos, como por ejemplo: Apache Cassandra, Neo4j, NebulaGraph y Memgraph, entre otros[14].
7. Investigar, desarrollar casos de estudio e incorporar nuevas pruebas para la escalabilidad horizontal de motores de Bases de Datos No Relacionales [9].
8. Repositorios para la sincronización del trabajo en equipo [19, 30].
9. Modelo de procesos para el despliegue / puesta en producción de sistemas de software [3,30].

RESULTADOS OBTENIDOS/ESPERADOS

Los resultados esperados/obtenidos se pueden resumir en:

- Capacitación continua de los miembros de las líneas de investigación.
- Estudio y análisis de procesos de diseño para almacenamiento no estructurado de datos y para Bases de Datos no relacionales.[1, 28]
- Análisis comparativo de aplicación de procesos de diseños de Bases de Datos NoSQL. Estudio experimental [31, 33]
- Estudio y análisis de distintos motores de Bases de Datos para el desarrollo de aplicaciones móviles [10, 15, 21].
- Estudio y análisis de Bases de Datos como servicio en la nube [29, 32].
- Análisis sobre la ejecución de consultas en diversos casos de estudio para distintos tipos de motores de Bases de Datos no relacionales [9, 14].
- Estudio y análisis sobre escalamiento horizontal para pruebas de performance entre distintos motores de Bases de Datos no relacionales, principalmente orientado a Grafos [13, 14, 28, 32].
- Definición de procesos de Gestión de Incidencias utilizando repositorios GIT [11, 12, 20, 19].
- Análisis de metodologías para la interoperabilidad de sistemas web y aplicaciones móviles [11, 12].

Algunas de las transferencias realizadas por el III-LIDI relacionadas con este proyecto, entre otras, son:

- Mantenimiento de la aplicación móvil para la comunidad de la Facultad de Informática de la UNLP.

2 Mantenimiento del Sistemas de congresos, utilizado por la RedUNCI y otras entidades.

3 Mantenimiento de Sistemas de Gestión Administrativa de Instituciones Universitarias.

[3] FORMACIÓN DE RECURSOS HUMANOS

Los integrantes de esta línea de investigación dirigen Tesinas de Grado y Tesis de Postgrado en la Facultad de Informática, y Becarios III-LIDI en temas relacionados con el proyecto. Además, participan en el dictado de asignaturas/cursos de grado y postgrado de la Facultad de Informática de la UNLP.

BIBLIOGRAFÍA

[21] Un estudio de procesos de diseño de bases de datos NoSQL. Marrero Luciano, Osowy Verena, Tesone Fernando, Thomas Pablo Javier, Corbalán Leonardo César, Fernández Sosa Juan Francisco, Pesado Patricia Mabel. XXVIII Congreso Argentino de Ciencias de la Computación (CACIC) (La Rioja, 3 al 6 de octubre de 2022). ISBN: 978-987-1364-31-2. Páginas: 404-414.

[22] Aspectos de ingeniería de software, bases de datos relacionales, no relacionales y como servicios en la nube para el desarrollo de sistemas de software híbridos. Marrero Luciano, Thomas Pablo Javier, Pasini Ariel Cristian, Bertone Rodolfo Alfredo, Ibañez Eduardo Javier, Aguirre Verónica, Panizzi Marisa Daniela, Olsowy Verena, Tesone Fernando, Pesado Patricia Mabel. XXIV

Workshop de Investigadores en Ciencias de la Computación (WICC 2022, Mendoza). ISBN: 978-987-48222-3-9. Páginas: 370-376.

[31] Ingeniería de Software. Un Enfoque Práctico. Novena Edición. Roger S. Pressman, Bruce R. Maxim. McGraw-Hill Interamericana Editores, S.A. 2021. ISBN: 978-1-4562-8772-6.

[32] Ingeniería de Software Teoría y Práctica. Pflieger Shari Lawrence. Pearson / Prentice Hall. 2002. ISBN: 9789879460719

[33] Ingeniería de Software. Novena Edición. Ian Sommerville. Addison Wesley / Pearson. 2011. ISBN: 978-607-32-0603-7

[34] Diseño Conceptual de Bases de Datos, un enfoque de entidades-interrelaciones. Carlo Batini, Stefano Ceri, Shamkant B. Navathe.. Addison-Wesley/Díaz de Santos. ISBN 0-201-60120-6 (1994).

[35] Model Driven Extraction of NoSQL Databases Schema: Case of MongoDB. Brahim, A.; Ferhat, R. and Zurfluh, G. (2019). In Proceedings of the 11th International Joint Conference on Knowledge Discovery. V.1: KDIR, ISBN 978-989-758-382-7, pages 145-154.

[36] *NoSQL*: modelos de datos y sistemas de gestión de bases de datos. XX Workshop de Investigadores en Ciencias de la Computación (WICC 2018). Universidad Nacional del Nordeste (UNNE). <http://sedici.unlp.edu.ar/handle/10915/67258>

[37] Análisis de performance en Bases de Datos NoSQL y Bases de Datos Relacionales. Luciano Marrero, Verena Olsowy, Fernando Tesone, Pablo Thomas, Lisandro Delia y Patricia Pesado. XXVI Congreso Argentino

de Ciencias de la Computación (CACIC 2020). Universidad Nacional de La Matanza, del 5 al 9 de octubre del 2020.
<https://cacic2020.unlam.edu.ar/es-ar/>

10. Un Análisis Experimental de Sistemas de Gestión de Bases de Datos para Dispositivos Móviles. 2021. Tesone Fernando, Thomas Pablo, Marrero Luciano, Olsow Verena, Pesado Patricia. XXVII Congreso Argentino de Ciencias de la Computación (CACIC) (Modalidad virtual, 4 al 8 de octubre de 2021). Universidad Nacional de Salta (UNSA), Argentina.
<http://sedici.unlp.edu.ar/handle/10915/130353>
11. Aspectos de ingeniería de software, bases de datos relacionales y bases de datos no relacionales para el desarrollo de sistemas de software en escenarios híbridos. Luciano Marrero, Pablo Thomas, Ariel Pasini, Rodolfo Bertone, Eduardo Ibáñez, Verónica Aguirre, Verena Olsow, Fernando Tesone, Patricia Pesado. XXII Workshop de Investigadores en Ciencias de la Computación (WICC 2020). Universidad Nacional de la Patagonia Austral (UNPA). El Calafate, Santa Cruz (Mayo 2020).
<http://sedici.unlp.edu.ar/handle/10915/104026>.
12. Aspectos de ingeniería de software, bases de datos relacionales, y bases de datos no relacionales y bases de datos como servicios en la nube para el desarrollo de software híbrido. Marrero Luciano, Thomas Pablo, Olsow Verena, Pasini Ariel, Bertone Rodolfo, Ibáñez Eduardo, Aguirre Verónica, Panizzi Marisa, Pesado Patricia. XXIII Workshop de Investigadores en Ciencias de la Computación (CACIC 2021, Chilecito, La Rioja).
13. Aspectos de ingeniería de software y bases de datos para el desarrollo de sistemas de software en escenarios híbridos. Patricia Pesado, Rodolfo Bertone, Pablo Thomas, Ariel Pasini, Luciano Marrero, Eduardo Ibáñez, Alejandra Rípodas, Verónica Aguirre, Verena Olsow, Fernando Tesone. XXI Workshop de Investigadores en Ciencias de la Computación (WICC 2019). Universidad Nacional de San Juan (UNSJ). (Abril 2019).
<http://sedici.unlp.edu.ar/handle/10915/77088>.
14. Un estudio comparativo de bases de datos relacionales y bases de datos NoSQL. Pesado Patricia, Thomas Pablo, Delía Lisandro, Marrero Luciano, Olsow Verena, Tesone Fernando, Fernández Juan Sosa. XXV Congreso Argentino de Ciencias de la Computación (CACIC 2019). Universidad Nacional de Río Cuarto, Córdoba, 14 al 18 de octubre de 2019. ISBN 978-987-688-377-1.
<http://sedici.unlp.edu.ar/handle/10915/91403>.
15. A. Nori, "Mobile and embedded databases," in Proceedings of the 2007 ACM SIGMOD International Conference on Management of data, pp. 1175-1177, 2007.
16. SQLite. <https://www.sqlite.org/>. Accedido en febrero de 2024.
17. Couchbase. <https://www.couchbase.com/>. Accedido en febrero 2024.
18. Aplicaciones para Dispositivos Móviles. Estrategias y enfoques de desarrollo. Thomas Pablo Javier, et. al. XXI Workshop de Investigadores en Ciencias de la

Computación (San Juan 2019). ISBN: 978-987-3984-85-3

19. Which Change Sets in Git Repositories Are Related?. Ramadani, J., Wagner, S. EEE 2016. International Conference on Software Quality, Reliability and Security (Viena, Austria.)
20. Synchronization and replication in the context of mobile applications. STAGE, A. (2005 Joint Advanced Student School Course 6: Next-Generation User-Centered).
21. S. Lee, "Creating and using databases for android applications," International Journal of Database Theory and Application, vol. 5, no. 2, 2012.
22. III-LIDI: <https://weblidi.info.unlp.edu.ar/proyectos/investigacion/>
23. Firebase. <https://firebase.google.com/?hl=es>. Accedido en febrero de 2024.
24. Redis. <https://redis.io/>. Accedido en febrero de 2024.
25. MongoDB. <https://www.mongodb.com/es>. Accedido en febrero de 2024.
26. Apache Cassandra. https://cassandra.apache.org/_/index.html. Accedido en febrero de 2024.
27. Neo4j. <https://neo4j.com/>. Accedido en febrero de 2024.
28. Aplicación de Procesos de Diseño de Bases de Datos NoSQL. Luciano Marrero, Verena Olsowy, Fernando Tesone, Pablo Thomas, Patricia Pesado. XXIX Congreso Argentino de Ciencias de la Computación (CACIC 2023) (Lujan)
29. Performance Analysis in NoSQL Databases, Relational Databases and NoSQL Databases as a Service in the

Cloud. Luciano Marrero, Verena Olsowy, Fernando Tesone, Pablo Thomas, Lisandro Delia, Patricia Pesado. https://link.springer.com/chapter/10.1007/978-3-030-75836-3_11

30. Aspectos de ingeniería de software, bases de datos relacionales, no relacionales y como servicios en la nube para el desarrollo de sistemas de software híbridos. Marrero, Luciano; Thomas, Pablo Javier; Pasini, Ariel Cristian; Bertone, Rodolfo Alfredo; Ibañez, Eduardo Javier; Aguirre, Verónica; Panizzi, Marisa Daniela; Olsowy, Verena; Tesone, Fernando; Pesado, Patricia Mabel. XXIV Workshop de Investigadores de la Ciencia de la Computación. Mendoza, 2022.
31. NoSQL database design processes: a comparative study. Luciano Marrero, Verena Olsowy, Pablo Thomas. XI Jornadas de Cloud Computing, Big Data & Emerging Topics. La Plata, 2023.
32. Un estudio comparativo de Bases de Datos NoSQL. Luciano Marrero. <https://sedici.unlp.edu.ar/handle/10915/155904>
33. Un estudio de procesos de diseño de bases de datos NoSQL. Verena Olsowy. <https://sedici.unlp.edu.ar/handle/10915/159625>