

METODOLOGÍA PARA EVALUACIÓN DE IMPACTO DE MIGRACIÓN ENTRE VERSIONES DE BASES DE DATOS NoSQL

Giovanni Daián Róttoli^{1,2}, Juan Zaffaroni¹, Marcelo López Nocera¹, Ma. Florencia Pollo-Cattaneo^{1,2}

¹ Grupo de Estudio en Metodologías de Ingeniería de Software (GEMIS). Facultad Regional Buenos Aires. Universidad Tecnológica Nacional. Argentina.

² Ingeniería en Sistemas de Información, Facultad Regional Concepción del Uruguay. Universidad Tecnológica Nacional. Argentina
{gd.rottoli, flo.pollo}@gmail.com

RESUMEN

Día a día, el crecimiento sostenido de tecnologías como Internet y la proliferación de herramientas como las redes sociales disparan el surgimiento de necesidades en materia de almacenamiento y procesamiento de datos. Las distintas tecnologías NoSQL existentes, soluciones en desarrollo de las problemáticas asociadas al fenómeno Big Data, se actualizan constantemente para dar respuesta a esta amplia variedad de requerimientos. Sin embargo, estas nuevas versiones de los sistemas de base de datos NoSQL pueden dar lugar a cambios que impacten de manera diversa sobre los distintos procesos que hacen uso de los mismos. Por dicho motivo, en este trabajo se presenta la propuesta de elaboración de una metodología para la evaluación del impacto de la realización de migraciones entre distintas versiones de tecnologías NoSQL, a fin de otorgar herramientas que permitan a los usuarios evaluar alternativas y optimizar la toma de decisiones en lo que se refiere a la implementación de tales actualizaciones.

Palabras clave: NoSQL, Migración, Análisis de Impacto

CONTEXTO

La presente investigación se desarrolla dentro de una línea de trabajo incipiente del Grupo de Estudio de Metodologías para Ingeniería en Software y Sistemas de Información (GEMIS),

conformado por un equipo de docentes y alumnos de la Universidad Tecnológica Nacional (UTN), pertenecientes a las Regionales Buenos Aires y Concepción del Uruguay. El grupo GEMIS, se halla abocado a la búsqueda de la sistematización de cuerpos de conocimientos y promoción sobre el campo de la Ingeniería en Sistemas de Información y la Ingeniería en Software, sus aplicaciones y abordajes metodológicos en todo tipo de escenarios (convencionales y no convencionales).

En el marco de la UTN, el equipo de investigadores ha trabajado en forma conjunta desde los Departamentos de Ingeniería en Sistemas de Información de ambas Regionales y desde la Escuela de Posgrado de la Regional Buenos Aires, integrando entre sus miembros a docentes de grado y de posgrado, articulando los resultados de investigaciones con el desarrollo de Trabajos Finales de Carrera, Trabajos Finales de Especialidad y Tesis de Maestría.

1. INTRODUCCIÓN

De manera habitual las organizaciones mutan para adecuarse al impacto de los cambios del medio circundante. Dichas mutaciones involucran desde el aspecto humano hasta las

tecnologías de soporte a la toma de decisiones [Moya, 2008].

Las bases de datos, son una de las herramientas que deben adecuarse continuamente a los distintos y cambiantes requerimientos de los usuarios [Lopez, 2012]. Por tal motivo, en la última década, han surgido un conjunto de alternativas a las bases de datos tradicionales, que luchan contra las problemáticas acarreadas por el fenómeno del Big Data. Las denominadas bases de datos NoSQL han sido desarrolladas para brindar esquemas flexibles en el trabajo con datos no estructurados, alta eficiencia ante consultas de datos, y posibilidades de escalar horizontalmente, sacrificando cuestiones como la consistencia de los datos u otras características deseables en las bases de datos SQL [Abramova et al., 2014, Nayak et al., 2014, Strauch et al., 2011].

Bajo esta propuesta se establecen cuatro grandes tipos de tecnologías NoSQL [Arora & Aggarwal, 2013; Bugiotti & Cabibbo, 2013; Hecht & Jablonski, 2011, Moniruzzaman & Hossain, 2013; Nayak et al., 2013; Strauch et al., 2011]:

- **Documentales**, las cuales permiten almacenar información en forma de documentos sin un esquema predefinido, usualmente haciendo uso de archivos JSON o XML referenciados por claves únicas. Ejemplos de ello son: MongoDB [Mongodb.org, 2016], CouchDB [Couchdb.apache.org, 2016] y Riak [Basho, 2016].
- **Clave-Valor**, que resultan muy simples y eficientes. Asimilándose a mapas y diccionarios, almacenan información en dos partes, una parte “valor” (que

almacena información), y una parte “clave” (que referencia a dicho valor). Los registros no pueden relacionarse entre sí, sino en la lógica de la aplicación que las utiliza. Ejemplos de bases de datos Clave-Valor son Redis [Redis.io, 2016] y Voldemort [Project Voldemort, 2016].

- **Columnares**, permiten referenciar a partir de una clave un conjunto de valores o columnas que pueden, a su vez, agruparse en familias de columnas, admitiendo un nivel de organización de la información, superior al de las bases de datos clave-valor. Sin embargo, al igual que éstas, persisten en la imposibilidad de relacionar entidades. De esta manera operan bases de datos como Cassandra [Cassandra.apache.org, 2016] y Hbase [Hbase.apache.org, 2016].
- **Gráficas**, acceden al almacenamiento de la información en forma de grafos, a través de nodos y relaciones entre ellos. Estas bases de datos son muy eficientes para realizar consultas entre entidades altamente relacionadas (como relaciones de amistad entre entidades personas, o productos relacionados con operaciones de compra) pues se sirven de las aplicaciones de la Teoría de Grafos. La utilización de bases de datos gráficas puede observarse en Neo4J [Neo4j Graph Database, 2016].

Como se puede apreciar, cada tipo de base de datos NoSQL posee un conjunto de características que deben tenerse en cuenta a la hora de seleccionar la más adecuada para cada organización [Pollo-Cattaneo et al., 2014]. La

importancia de la valoración de los factores distintivos de cada tecnología se evidencia aún más, si se ponen en perspectiva el auge y la democratización del acceso a Internet y las telecomunicaciones, la impronta que el avance de éstas confiere al surgimiento de nuevas necesidades y la fuerza con la que imponen el desarrollo e implementación de funcionalidades novedosas que den respuesta a tales requerimientos emergentes; demandas que el mercado informático ha de cubrir, poniendo a disposición las más variadas características en cada nueva versión de la tecnología en cuestión [Stanier, 2012].

Sin embargo, en ocasiones, al elegir operar con versiones actualizadas, se sacrifican utilidades que los usuarios valoran por lo que es altamente conveniente que en la elección para la utilización de un motor de base de datos NoSQL, se considere de importancia no solo el tipo de base de datos, sino también, el impacto que ha de tener la migración desde una instancia del software hacia otra, desde el punto de vista de las implicaciones en los procesos que hacen uso de las mismas [Pollo-Cattaneo et al., 2014].

Ante esta situación, se hace necesaria una metodología que permita la comparación de versiones de bases de datos NoSQL a fin de poder evaluar el impacto de migración entre dichas versiones.

2. LÍNEAS DE INVESTIGACIÓN, DESARROLLO E INNOVACIÓN

Durante 2015, a pedido de una organización del sector privado, el grupo GEMIS realizó un análisis comparativo de dos versiones de MongoDB₇ (base de datos NoSQL de tipo Documental) en un ambiente empresarial, a fin de evaluar las características de ambas, el

impacto que provocan sobre los procesos productivos relacionados y la migración de la tecnología utilizada hacia una versión más actualizada de la misma [Zaffaroni et al., 2016].

Frente a tal escenario, se evidenció la necesidad de contar con una serie de pasos sistematizados a efectos de replicar la evaluación efectuada, sobre cualquier tipo de base de datos no relacional, y obtener precisiones respecto de las cualidades más relevantes de las tecnologías implicadas, y de un posible impacto sobre los procesos productivos que se ponen en juego en las instancias de actualización.

En este contexto, el objetivo que se persigue es el de desarrollar una metodología de trabajo que pueda aplicarse entre versiones de cualquier tipo bases de datos NoSQL y que permita a sus usuarios acceder a herramientas para la selección eficiente de la más adecuada. (La metodología propuesta será evaluada ante casos concretos a fin de asegurar su validez).

Para cumplir con el objetivo establecido, se seguirán los siguientes pasos:

1. Realizar una investigación documental exploratoria sobre las características relevantes de las bases de datos NoSQL.
2. Identificar casos y datos para las futuras pruebas.
3. Seleccionar las condiciones sobre las cuales se ejecutarán los casos de prueba y validación.
4. Desarrollar una propuesta inicial de la metodología.
5. Estudiar el comportamiento del modelo utilizando los casos de prueba y validación sobre los datos propuestos.

6. Analizar los resultados obtenidos y realizar modificaciones hasta contar con una versión estable.

3. RESULTADOS Y OBJETIVOS ESPERADOS

Mediante la ejecución del procedimiento descrito en el apartado anterior, se busca desarrollar una metodología de trabajo para la comparación de versiones de bases de datos NoSQL a fin de evaluar los posibles impactos de migración entre ambas tecnologías.

Dicha metodología permitirá conocer si ciertas características presentes en determinadas versiones presentan cambios y el tipo de cambio en cuestión (actualización o eliminación de la característica, entre otros). De esta manera, se podrá comparar si un cambio de versión en el sistema de bases de datos NoSQL afectaría al sistema del cual forma parte.

El proyecto propuesto se inscribe en una línea de investigación dentro del ámbito educativo universitario mediante el estudio y análisis de diferentes tecnologías de Bases de Datos relacionales y no relacionales. Con ello, se pretende continuar y ampliar los trabajos desarrollados por GEMIS en el área de Bases de Datos, brindando conocimiento valioso para volcar en las aulas de grado y posgrado y, en una segunda instancia, realizando difusión de los mismos dentro del ámbito científico.

4. FORMACIÓN DE RECURSOS HUMANOS

Este proyecto aspira, tanto a la obtención de nuevos conocimientos y a la formación de especialistas en el análisis de la adopción de procesos vinculados con la Ingeniería de Software, como a la promoción del ascenso de

los recursos humanos implicados, dentro del escalafón de la carrera de investigadores. El grupo de trabajo está integrado por dos investigadores formados, un investigador en formación y un becario doctoral. Además, uno de los miembros se encuentra en desarrollo de su Trabajo Final de Especialidad.

Asimismo, en el marco de este proyecto de investigación se prevé la radicación de una Tesis de Maestría en Ingeniería en Sistemas de Información y su articulación con trabajos finales de la carrera de grado de Ingeniería en Sistemas de Información.

5. REFERENCIAS

- Abramova, V., Bernardino, J., & Furtado, P. (2014). Experimental evaluation of NoSQL databases. *International Journal of Database Management Systems*, 6(3), 1.
- Arora, R., & Aggarwal, R. R. (2013). Modeling and Querying Data in MongoDB. *International Journal of Scientific and Engineering Research*, 4(7).
- Basho, (2016). Riak KV. Retrieved 29 February 2016, from <http://basho.com/products/riak-kv/>
- Bugiotti, F., & Cabibbo, L. (2013). A Comparison of Data Models and APIs of NoSQL Datastores. Dipartimento di Ingegneria della Università di Roma.
- Cassandra.apache.org. (2016). The Apache Cassandra Project. Retrieved 29 February 2016, from <http://cassandra.apache.org/>
- Couchdb.apache.org. (2016). Apache CouchDB. Retrieved 29 February 2016, from <http://couchdb.apache.org/>
- Hbase.apache.org. (2016). Apache HBase – Apache HBase™ Home. Retrieved 29 February 2016, from <https://hbase.apache.org/>
- Hecht, R., & Jablonski, S. (2011). NoSQL evaluation: A use case oriented survey. 2011 International Conference on Cloud and Service Computing. 336-341.
- Lopez, D. (2012). Análisis de las posibilidades de uso de Big Data en las organizaciones. Universidad de Cantabria, Santander, España.
- Mongodb.org. (2016). MongoDB. Retrieved 29 February 2016, from <https://www.mongodb.org/>
- Moniruzzaman, A. B. M., & Hossain, S. A. (2013). Nosql database: New era of databases for big data analytics-classification, characteristics and comparison. arXiv preprint arXiv:1307.0191.

- Moya J. (2008). *Management Democrático*. Cataluña: PreMya Consultores.
- Nayak, A., Poriya, A., & Poojary, D. (2013). Type of NOSQL databases and its comparison with relational databases. *International Journal of Applied Information Systems*, 5(4), 16-19.
- Neo4j Graph Database. (2016). Neo4j: The World's Leading Graph Database. Retrieved 29 February 2016, from <http://neo4j.com/>
- Pollo-Cattaneo, M. F., Nocera, M. L., & Rottoli, G. D. (2014). Rendimiento de tecnologías NoSQL sobre cantidades masivas de datos. *Cuaderno Activa*, (6), 11-17. ISSN: 2027 – 8101.
- Project Voldemort,. (2016). Project Voldemort: A Distributed Database. Retrieved 29 February 2016, from <http://www.project-voldemort.com/voldemort/>
- Redis.io,. (2016). Redis. Retrieved 29 February 2016, from <http://redis.io/>
- Stanier, C. (2012). Introducing nosql into the database curriculum. In 10th International Workshop on the Teaching, Learning and Assessment of Databases (p. 61).
- Strauch, C., Sites, U. L. S., & Kriha, W. (2011). NoSQL databases. Lecture Notes, Stuttgart Media University.
- Zaffaroni, J., Róttoli, G.D., López Nocera, M., Pollo-Cattaneo, M.F., (2016). Reporte Técnico: Estudio Comparativo de características de versiones de la base de datos NoSQL Documental MongoDB. Retrieved 02 March 2016, from <https://sistemas.frba.utn.edu.ar/grupogemis/Trabajos/Reportes%20Tecnicos/GEMIS-TD-2016-03-TR-2016-03.pdf>