

Semánticas para las Buenas Prácticas en el Diseño de Datos en Bases de Datos Relacionales

Muñoz Roberto Miguel; Damiano Luis Esteban; Maldonado Calixto; Romero María Soledad; Quinteros Sergio Ramón; Bueno Matías; Peretti Juan; Guevara Andrea

Departamento Ingeniería en Sistemas de Información
Facultad Regional Córdoba / Universidad Tecnológica Nacional
Maestro Marcelo López esq. Cruz Roja Argentina – Ciudad Universitaria - Córdoba
0351 – 4686385

{robertmunioz, luis.damiano, calixtomaldonado, romeroma.soledad, ser.quinteros, matiasbueno, peretti.juan, andrezza77}@gmail.com

Resumen

El proyecto de investigación, que da origen a este artículo, tiene entre sus metas lograr una coherencia discursiva en el ambiente áulico, para el dictado de clases en asignaturas que imparten conocimientos de base de datos relacionales. La propuesta es generar un conjunto de normas y reglas de buenas prácticas que den luz en temas que se tratan habitualmente y que no se encuentran taxativamente delineados en la bibliografía actual. El equipo investigador ha observado que en el espacio temático normalización de modelos de datos no hay suficiente claridad y definiciones respecto del desarrollo de la estructura de datos. Surgen entonces algunos interrogantes que no tienen respuesta con fundamento bibliográfico: ¿Cuántas soluciones de normalización (modelo conceptual) se pueden alcanzar para un problema de almacenamiento de datos? ¿Cuáles son las formas aconsejables de construcción de un modelo de datos? Por ello la hipótesis de investigación planteada es “el diseño de un modelo conceptual de base de datos relacional, para un problema determinado y acotado, tiene una solución única”. Para trabajar en este proyecto, el equipo considera imprescindible la participación de docentes y profesionales informáticos, que se desempeñan en distintos ámbitos y con distintas competencias, para validar o refutar la hipótesis.

Palabras clave: *normalización, modelo relacional, diseño bases de datos.*

1. Identificación

Código del PID: “SIUTNCO0004979”.

Título: “Buenas prácticas en el diseño de estructuras de datos en bases de datos relacionales”.

Tema prioritario del Programa de Tecnología Educativa y Enseñanza de la Ingeniería: “La enseñanza de la ingeniería y la formación de los ingenieros”.

Fecha de inicio: 01/01/2018.

Fecha de finalización: 31/12/2019.

2. Introducción

Este proyecto homologado por la Universidad Tecnológica Nacional (UTN) surge en el seno de la cátedra Gestión de Datos (GDA), de tercer año de la carrera de Ingeniería en Sistemas de Información de la Facultad Regional Córdoba (FRC). Está conformada por más de diez docentes, donde se suelen suscitar discusiones referidas a los resultados o soluciones de diseño de bases de datos, sin lograr una postura generalizada.

Los expertos en base de datos, de renombre global y local, han publicado una gran variedad de libros que dan fundamentos sobre el tema que se plantea en el proyecto.

La bibliografía, incluso la considerada obligatoria en la cátedra, describe en detalle al modelo relacional, respecto a: estructura del modelo, integridad, manipulación de datos, alcances y limitaciones, pero no se

encontró respuesta en ninguno de ellos al interrogante: Ante un requerimiento de diseño de base de datos relacional y considerando que es provista por distintos arquitectos de datos, ¿puede ser única la estructura de datos normalizada? Esta pregunta puede merecer mayor atención con el fin de generar conocimiento. La normalización, como técnica para crear estructuras de bases de datos relacionales en estado óptimo y deseable (normalizado), evitando redundancia de datos y futuras inconsistencias, ha sido abordada por múltiples autores como C.J. Date en su libro “Introducción a los Sistemas de Base de Datos” [1], donde ha enfocado su atención en aspectos técnicos que marcan una tendencia clara en el conocimiento y opinión de los arquitectos de datos. «...En otras palabras, "normalizada" y "1FN" significan exactamente lo mismo; aunque debe tener presente que el término "normalizado" se usa a menudo para indicar uno de los niveles más altos de normalización (por lo regular a la tercera forma normal, 3FN)...» [1] (Página 349). Ramez Elmasri y Shamkant B. Navathe expresan en su libro “Fundamentos de Sistemas de Bases de Datos” [2] «... Así, el diseño de base de datos, tal y como se realiza en la actualidad en la industria, presta especial atención a la normalización hasta la 3FN, la BCNF o la 4FN...» [2] (Página 300). Ambos autores expresan una aproximación a los límites prácticos del concepto “Normalización”, desde lo observado en el campo de la Industria del Software. No hay una estricta coincidencia en el alcance, pero sí expresan una conceptualización con algún grado de aproximación, en función del criterio de cada uno. En ambos casos los autores coinciden que se logra la Normalización a partir de la 3FN, como mínimo. A su vez Silberschatz trata el concepto Normalización con la misma visión de Date y Elmasri, en el capítulo 7 de su libro “Fundamentos de Bases de Datos” [3]. Cabe destacar que se permite expresar un singular punto de vista sobre la Normalización que expresa: «...Cuando se define con cuidado un diagrama E-R, identificando correctamente todas las entidades, las tablas generadas a partir del

diagrama E-R no necesitan más normalización. No obstante, puede haber dependencias funcionales entre los atributos de una entidad...» [3] (Página 183), y en la misma página también expresa «... De manera alternativa, la normalización puede dejarse a la intuición del diseñador durante el modelado E-R, y puede hacerse formalmente sobre las relaciones generadas a partir del modelo E-R...». En el material de Silberschatz no se percibe que acentúe su mirada pragmática desde la observación de la industria informática sobre el tema normalización, como lo hicieron los otros dos autores citados y que les permitió expresar que la Normalización de una estructura de datos requiere que ésta alcance al menos la 3FN (tercera forma normal). Cuando se permite expresar su singular punto de vista, respecto de la Normalización, entiende que los aspectos intuitivos de un arquitecto de datos pueden llevarlo al diseño Normalizado de una estructura de datos.

Este tipo de conceptos, vertidos por los autores Date y Elmasri, han sido asimilados como una pieza de conocimiento que no tiene discrepancia por la gran mayoría de los especialistas que diseñan estructuras de datos, por lo que no existen controversias.

Sin embargo, no se pudo encontrar aún, entre los autores estudiados, que alguno de ellos haya orientado su análisis y explicación de la Normalización sobre seguir una metodología que permita realizar esta acción, demostrando que la estructura resultante es única, partiendo de una relación universal (entidad primera que reúne todos los atributos observados) como lo expresa Silberschatz, en el “El enfoque de la relación universal” en el libro ya citado «... El segundo enfoque del diseño de bases de datos es comenzar con un solo esquema de relación que contenga todos los atributos de interés y descomponerlo...» [3] (Página 183).

El grupo de investigación propone, en función de su experiencia de campo y en la docencia, demostrar que el modelo conceptual para un problema de almacenamiento de datos es único, en el marco de base de datos relacionales y

cumpliendo con las reglas de la normalización. Dicho de otra manera, dada una situación, con requerimientos a igual nivel de detalle, el conjunto solución de relaciones normalizado al que arriban distintos grupos de trabajo es único.

Se ha realizado una intensa búsqueda bibliográfica para encontrar opiniones, discusiones, afirmaciones o negaciones del problema planteado. En dicha búsqueda, en los materiales abordados, detallados en referencia bibliográfica, no se hallaron resultados que conformen un antecedente.

Las búsquedas también incluyeron anales de congresos Argentinos (WICC, CoNaISI, JAIIO) y sitios de internet, sin hallar resultados favorables sobre la temática específica.

Hay que destacar que existe unanimidad conceptual en el material de todos los autores, sobre las definiciones referidas al modelo relacional. Algunos con mayor énfasis en aspectos de principios, otros con más desarrollo en las características de las formas normales, pero en términos generales sin desacuerdos importantes.

Integrantes del equipo de investigación, a su vez, son docentes de la cátedra, donde trabajan los conceptos y aplicaciones de las bases de datos, que conforman los contenidos mínimos de la asignatura y constituyen el eje central de estudio. Los conocimientos son homogéneos entre los integrantes, sobre los conceptos inherentes a base de datos, modelo relacional y normalización de una estructura de datos. Esta tarea docente ha sido desempeñada durante más de veinte años, trabajando con estos conocimientos. Han presentado diversos artículos en congresos de alcance nacional, sobre distintas temáticas que involucran estos saberes. Esta participación involucra la actitud atenta al análisis del tema de hipótesis, siendo un factor importante para afirmar que el problema no está siendo tratado en este momento y no presenta antecedentes expuestos en el pasado.

3. Objetivos, Avances y Resultados

El objetivo general del proyecto es: proponer un conjunto de reglas de buenas prácticas para el diseño del modelo conceptual de datos, normalizado, para bases de datos relacionales, tendientes a explicar las restricciones y características que fundamentan que el modelo conceptual de una base de datos relacional puede ser único, para un dominio dado.

A partir del objetivo seguido, el material a generar desde el proyecto aportará documentación de referencia, dando soporte a la actividad áulica en la temática. Se pretende elaborar guías que colaboren en:

- Identificar y caracterizar las estructuras típicas que se utilizan en una solución.

- Identificar cuando aplique, diferentes estructuras que pueden ser equivalentes, resaltando similitudes, ventajas y diferencias de esas estructuras.

- Entender y explicar la forma en que implementan una estructura de datos válida los arquitectos de datos.

- Sugerir un método de normalización, en la búsqueda de estandarizar el proceso.

- Definir y aplicar reglas de buenas prácticas.

Actualmente el equipo ha avanzado en la recolección de información a través de encuestas y experimentos para demostrar o refutar la hipótesis. La mirada se enfoca en una conclusión final que surja del análisis de toda la información obtenida en el transcurso de los años de investigación. Es prematuro en este momento poder dar una tendencia u opinión de los temas tratados, visto que hace poco tiempo que se están recolectando datos y experiencias. Las encuestas están direccionadas a observar el pensamiento y forma en que un arquitecto de datos resuelve un problema de estructuras datos.

También se trabaja en caracterizar las estructuras que intervienen en las soluciones de diseño de bases de datos. En este punto se ha detectado que las mismas son un número reducido de opciones, más allá que se sigue indagando, con la particularidad que ninguna estructura compite con otra, en relación al desempeño en el resultado. En otras palabras, una estructura no tiene otra

estructura que la sustituya. En este tema de caracterizar estructuras el grupo de trabajo ya cuenta con un camino recorrido, en la actividad profesional y en el trabajo de cátedra. Esto permite vislumbrar una tendencia, pero el propósito es continuar consultando y encuestando a profesionales. Se considera un avance significativo poder establecer claridad conceptual en torno del problema de la hipótesis, pues en el ámbito de la docencia brindará una posición clara el poder afirmar o negar la hipótesis planteada. También subyace en forma encubierta una hipótesis derivada de la propuesta en el proyecto de investigación: “Una solución de diseño de estructuras de datos, para base de datos relacionales y para un mismo problema, es distinta a otra si la interpretación del problema, por parte de los arquitectos de datos de datos, también lo es”. Esta hipótesis aún no declarada encontraría respuesta con los resultados de esta investigación.

4. Formación de Recursos Humanos

El equipo de investigación está compuesto por: un director, un co-director, cuatro investigadores de apoyo, cuatro técnicos de apoyo y tres estudiantes becarios. Cabe aclarar que ocho de los integrantes se desempeñan como docentes en la FRC y en particular en la asignatura GDA.

El resultado final de este proyecto de investigación será un documento que plasme los conocimientos logrados. La principal utilidad consistirá en señalar un camino recomendado de buenas prácticas en la construcción de estructuras de datos para base de datos relacionales; incluyendo la demostración o rechazo de la hipótesis, en virtud de las evidencias prácticas que se hayan logrado recolectar. Si se logra demostrar que para un problema acotado existe un conjunto determinado de estructuras, iguales en todas las soluciones observadas y que conforman una solución satisfactoria al problema de almacenamiento de datos, la demostración conformará una caracterización consensuada (ontológica).

La documentación a generar tendrá toda la información recogida, con la ejemplificación

explicativa, dando sustento a las explicaciones que devengan de la investigación. Se confía en que el producto obtenido será de aprovechamiento en el medio educativo. Brindará una metodología para la construcción de una estructura de datos, sugiriendo cuales deben ser los principios de buenas prácticas para este ámbito de la informática.

5. Publicaciones relacionadas con el PID

Algunas publicaciones relacionadas, donde integrantes del equipo han participado, son:

- Paz Menvielle Alejandra, et al. (2015) *Desarrollo de un Sistema Experto para el Proyecto de Investigación “Generador Automático de Modelos de Datos Normalizados en Bases de Datos Relacionales”*. Workshop de Investigadores en Ciencia de la Computación (WICC).

- Paz Menvielle, María Alejandra et al. (2014) *Una Estrategia de Aprendizaje Autónomo para Sistemas Expertos, con Aplicación Concreta al Proyecto – Generador Automático de Modelos de Datos Normalizados en Bases de Datos Relacionales*. 2do. Congreso Nacional de Ingeniería en Informática / Sistemas de Información (CoNaIISI).

- Muñoz Roberto Miguel, Maldonado Calixto, Damiano Luis Esteban, Romero María Soledad, Bueno Matías, Quinteros Sergio Ramón, Guevara Andrea, Peretti Juan Pablo, Carrasco Agustín (2018) *Buenas Prácticas En El Diseño De Estructuras De Datos En Bases De Datos Relacionales*. Workshop de Investigadores en Ciencia de la Computación (WICC).

- Roberto Miguel Muñoz, Calixto Maldonado, Luis Esteban Damiano, María Soledad Romero, Sergio Ramón Quinteros, Andrea Guevara y Agustín Carrasco (2017) *Implementación de la Plataforma Web MultiDB para enseñanza y aprendizaje del Lenguaje SQL*. V Jornada de Enseñanza de la Ingeniería (JEIN)

- Roberto Miguel Muñoz, Luis Esteban Damiano, Calixto Maldonado, María Soledad Romero, Sergio Ramón Quinteros, Andrea Guevara, Agustín Carrasco (2017) *Implementación de Plataforma de*

Aprendizaje del Lenguaje SQL. 5to. Congreso Nacional de Ingeniería en Informática / Sistemas de Información (CoNaIISI).

- Cuevas Juan Carlos et al. (2016) *Plataforma Web para acceder a diferentes Sistemas de Gestión de Bases de Datos Relacionales*. Workshop de Investigadores en Ciencia de la Computación (WICC)

- Reinosa Enrique; Maldonado Calixto; Muñoz Roberto; Damiano Luis; Abrutsky Maximiliano (2012) *Libro Bases de Datos*. Editorial AlfaOmega.

El grupo transmitirá los resultados del proyecto en un documento de dominio público que será puesto a disposición de las unidades académicas donde se imparten conocimientos de bases de datos.

Es intención del equipo investigador también publicar en formato de libro los resultados de la investigación, afirmando el resultado obtenido, tanto por la verificación o no de la hipótesis.

Referencias

[1] Date, Christopher. (1993) *Introducción a los Sistemas de Bases de Datos Volumen 1* - Quinta Edición - Edit.: Addison Wesley Iberoamericana

[2] Elmasri, R. y Navathe, S. (2007) *Fundamentos de Sistemas de Bases de Datos-5ta Edición* – Edit.: Pearson

[3] Silberschatz et al. (2006) *Fundamentos de Bases de Datos*. Quinta Edición- Edit. Mc Graw Hill

[4] Reinosa, E. et al. (2012) *Bases de Datos* - Edit.: AlfaOmega Editores

Otro material consultado:

-Abraham Silberschatz Bell Laboratories et al. (2007) *Fundamentos de Bases de Datos*. Cuarta edición – Edit.: Mcgraw-hill/Interamericana

-Becerra Zepeda, Sergio Antonio (1999) *Bases de Datos Inteligentes- Tesis para Maestro en Ciencias Área Computación* - Edit.: Universidad de Colima, México http://digeset.uco.mx/tesis_posgrado/Pdf/Sergio%20Antonio%20Becerra%20Zepeda.pdf. Accedido en febrero 2012.

-Dolores Costal Costa et al. (2007) *Databases* – Edit.: Universitat Oberta de Catalunya

http://ocw.uoc.edu/computer-science-technology-and-multimedia/bases-de-datos/Course_liseng (sept.2015)

-Mario Bunge (2011) *Tratado de Filosofía Ontología I* – Edit.: Gedisa

-Mario Bunge (2011) *Tratado de Filosofía Ontología II* – Edit.: Gedisa

-Mendelzon, Ale (2000) *Introducción a las Base de Datos Relacionales*. Edit.: Pearson

-Mercedes Marqués (2011) *Bases de Datos* - Edit.: Publicacions de la Universitat Jaume I. Servei de Comunicació i Publicacions.

-Neeraj Sharma et al. (2010) *Database Fundamentals Ideal for application developers and administrator* – Editorial: Copyright IBM

-Rafael Camps Paré et al. (2007) *Software Libre Base de Datos* – Edit.: Fundació per a la Universitat Oberta de Catalunya