

BUENAS PRÁCTICAS EN EL DISEÑO DE ESTRUCTURAS DE DATOS EN BASES DE DATOS RELACIONALES

Muñoz Roberto Miguel, Maldonado Calixto, Damiano Luis Esteban, Romero María Soledad, Bueno Matías, Quinteros Sergio Ramón, Guevara Andrea, Peretti Juan Pablo, Carrasco Agustín, Urbano Barbara, Arguello Santiago

Centro de Investigación, Desarrollo y Transferencia de Sistemas de Información (CIDS)
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, calixtomaldonado, luis.damiano, romeroma.soledad, matiasbueno, ser.quinteros, andrezza77, peretti.juan, asermax, sbaarby}@gmail.com

RESUMEN

El propósito de este proyecto de investigación es desarrollar un conjunto de reglas de buenas prácticas para el diseño de modelos conceptuales para bases de datos relacionales.

El grupo de trabajo parte de una idea expresada en el libro Bases de Datos [1] y que en este proyecto fue transformada en hipótesis. La hipótesis de trabajo es que el diseño de un modelo conceptual de base de datos relacional, para un problema determinado y acotado, tiene una solución única, la cual puede ser validada en el resto del artículo.

Están planificadas las actividades y acciones que permitan recolectar información para analizar la problemática, con instrumentos tales como: encuestas, entrevistas personales y grupales y experimentos, incluyendo el análisis cuantitativo y cualitativo de los datos.

Con este proyecto de investigación se logrará como resultado final un documento que exprese los resultados de la investigación, que afirmen o nieguen la hipótesis. Es objetivo es obtener un compendio de reglas y principios de buenas prácticas, en el diseño del modelo conceptual de bases de datos relacionales. El grupo de investigación aspira a darle continuidad en el tiempo, publicarlo en medios especializados, incluirlo en capítulos en libros de la temática, darle visibilidad en reuniones científicas y

presentarlo públicamente en el ámbito académico.

Palabras clave: Bases de datos, modelo relacional, arquitecto de datos, estructura de datos, diseño de bases de datos.

CONTEXTO

El proyecto ha sido homologado por la Secretaría de Ciencia, Tecnología y Posgrado de la Universidad Tecnológica Nacional, con un tiempo de realización de 24 meses, desde el 1 de enero de 2018. La Unidad Científico-Tecnológica donde se desarrolla es el Centro de Investigación, Desarrollo y transferencia de Sistemas de Información (CIDS), de la Universidad Tecnológica Nacional, Facultad Regional Córdoba (UTN-FRC).

El origen del proyecto se centra en observaciones realizadas por los docentes de la cátedra Gestión de Datos, de la carrera de Ingeniería en Sistemas de Información, de la UTN-FRC. En este ámbito, y también en el desarrollo profesional de ellos, se pudo ver que los modelos conceptuales, solución de un problema de datos, tenían una mirada homogénea por parte de quienes discutían la solución, encontrando pocas o ninguna discrepancia.

Cabe aclarar que los docentes también se desempeñan en cátedras relacionadas a las siguientes temáticas: Programación, Bases de

Datos, Dirección de Proyectos e Integración de Sistemas de Información.

El equipo ha trabajado de manera continua en proyectos de investigación, relacionados a la temática de bases de datos, motores de bases de datos y específicamente en diseño de estructuras de datos:

- TecnoDB - Administrador de Base de Datos Relacional (2007) [2]
- PROMETEO - Desarrollo de un método y una herramienta para el aprovechamiento de Metadatos de Base de Datos Relacionales (2010) [3]
- Análisis y aplicación de metodologías para la generación de consultas complejas utilizando esquemas OLAP (2010) [4]
- Generador Automático de Modelos de Datos Normalizados en Bases de Datos Relacionales (2014-2015) [5]
- MultiDB. Plataforma Web para acceder a diferentes Sistemas de Gestión de Bases de Datos Relacionales (2016-2017) [6]

1. INTRODUCCIÓN

El modelo relacional es una forma de estructuración de datos, para su almacenamiento y manipulación. Abarca distintos estadios que van desde la comprensión de la solución de un problema de almacenamiento, pasando por el modelo ontológico de éste (habitualmente denominado modelo conceptual), hasta la implementación de esta solución a través de un motor de base de datos.

El arquitecto de la base de datos, para entender un problema de gestión y almacenamiento, debe en primer lugar poder interactuar con los expertos en el dominio del problema, que generalmente no son especialistas en la tarea de modelar estructuras de datos. Dichos expertos, en el análisis, diseño y dominio del problema, suelen ser uno o varios profesionales involucrados con el flujo de datos y sus características, sin ser necesariamente profesionales con conocimiento formal sobre el diseño de bases de datos. El arquitecto tiene que ser una persona con amplitud en su forma de

compresión, en la lectura de documentación relacionada al problema que está abordando, pues ésta puede ser fundamental para la identificación de aspectos procedimentales, restricciones técnicas y otros limitantes que puedan surgir de la misma.

En términos generales un arquitecto de base de datos debe poder aprender del problema que está tratando tanto como pueda, a fin de alcanzar una comprensión clara del dominio del problema, como expresa Giarratano en su libro *Sistemas Expertos...*[7], en página 4, “el dominio del conocimiento en el experto, debe crecer hasta comprender el dominio del problema”.

Los autores del libro *Base de Datos* [1], en la página 69 indicaron: “Es conveniente que se tenga una idea muy clara sobre el dominio del problema y cuál es su límite, pues esto permitirá que se alcance un resultado definido y limitado que se manejará con facilidad”. De esta manera, al especificar un límite concreto, claro y conceptualmente definido, se separa el estricto interés que tratará la solución del conocimiento general y extenso, de cualquier área a la que pertenezca el problema. En otros términos es una zona acotada, con un límite visible e invariable, que es una restricción del conocimiento general a su área de pertenencia. Es de vital importancia que el arquitecto conceptualice el modelo desde un entendimiento y comprensión, y no desde una interpretación. Esta última puede alejarlo de la realidad del problema tratado.

El arquitecto, habiendo recabado suficiente información del equipo de análisis y diseño sobre el problema que está tratando, será capaz de realizar su tarea con experticia, desarrollando el diseño de la estructura de datos (especificación de la conceptualización).

2. LÍNEA DE INVESTIGACIÓN, DESARROLLO

La conceptualización de la solución sobre un problema analizado, que se representa a través de un esquema (gráfico), compuesto por entidades, atributos y relaciones; habitualmente

se llama esquema conceptual o modelo conceptual.

Las entidades se identifican por un nombre, que es consecuencia de la reunión de un conjunto de atributos únicos en ésta, que la caracterizan y califican. Una entidad es única dentro de un modelo conceptual o estructura de datos resultado. Mantienen un estado de asociación entre sí, que están representados por la vinculación que se produce entre las entidades a través de las claves foráneas y primarias. Es imperativo que el estado de asociación esté correctamente planteado, pues es fundamental para asegurar la integridad de los datos en la estructura diseñada.

La clave primaria dentro de cada entidad está compuesta por un atributo, o un conjunto de estos, que producen la identificación de una tupla en forma unívoca e inequívoca dentro de la entidad. Esta clave primaria, que debe cumplir con las propiedades de unicidad y minimalidad [8], tiene una relación unívoca con el conjunto de atributos no clave que representa.

Para que una entidad quede relacionada a otra se debe establecer la vinculación, incorporando dentro de ésta los atributos de la clave primaria de la entidad con la que se desea vincular. De esta manera se identifica a un conjunto de atributos que recibe la denominación de clave foránea o ajena.

Definiendo las entidades con su clave primaria, y su estado de asociación con las claves foráneas, se obtiene el modelo conceptual - lógico del resultado terminado.

Un arquitecto de datos puede emplear cualquier procedimiento para determinar la estructura de datos, no hay uno definido. Suelen utilizarse metodologías: ascendente, denominada diseño por síntesis, o descendente, diseño por análisis [9]. Con la experiencia termina elaborando su proceso/receta.

Académicamente el grupo de docentes basa la enseñanza de estos temas en aplicar la técnica de normalización, con el análisis de dependencias funcionales y la aplicación de las reglas definidas como formas normales por

Codd [8], con un proceso “que sigue un método descendente evaluando cada relación contra el criterio de las formas normales y descomponiendo las relaciones según sea necesario” [9]. La estructura de datos lograda se menciona según el nivel alcanzado, por ejemplo: Normalización a tercera forma normal, normalización a segunda forma normal, etc. Este estado alcanzado por la solución, está vinculado a la entidad que alcanzó la forma normal menos avanzada de nivel dentro de la estructura resultado. Por ejemplo: para que una estructura de datos tenga la calificación de normalizada a tercera forma normal, todas las entidades componentes de la estructura deben estar en tercera forma normal. Si una de ellas se encuentra en segunda forma normal el resultado total está en segunda forma normal. Como ya se mencionó, la bibliografía avala y expresa los aspectos teóricos del diseño de bases de datos, con mayor o menor fluidez.

Como consecuencia de las prácticas en el aula y el ejercicio profesional de los integrantes, en el diseño de estructuras de datos, se ha observado que las posibilidades de desarrollar estructuras equivalentes (dos o más estructuras distintas) que proporcionen el mismo alcance de resultado no han sido demostradas hasta este momento.

En el año 2010 un grupo de docentes de la Universidad Tecnológica Nacional (Ing. Roberto Muñoz, Ing. Calixto Maldonado, Lic. Luis Damiano, Ing. Enrique Reinoso y el Ing. Maximiliano Abrutsky) escribieron un libro, para la editorial AlfaOmega, con el título: Base de Datos [1] y en él se presentó el pensamiento relacionado con esta afirmación que se acaba de esbozar. Este libro fue sometido, por parte de la editorial, a la consideración de distintos especialistas en bases de datos, de universidades locales y también de latinoamérica, y todos coincidieron que la afirmación no tenía precedente bibliográfico y admitieron aceptar el concepto con la siguiente salvedad en la página 67 [1]: “... Si bien el pasaje de un conjunto de relaciones no normalizadas a un conjunto de relaciones se

puede realizar de cualquier forma, el resultado debe ser siempre el mismo, independientemente del método que se haya utilizado. La bibliografía consultada en la elaboración de este capítulo, difiere con este autor en la manera de abordar el concepto “normalización”. Sin embargo, la metodología aquí utilizada se considera indispensable para la comprensión total del tema.”

Por todo esto, el grupo de investigación propone la discusión respecto a la siguiente hipótesis a demostrar “Si distintos arquitectos de bases de datos:

- a) parten del mismo origen de datos y los entienden como la misma colección de datos,
- b) observan un problema de almacenamiento de datos inicial, por ejemplo, una planilla que contiene el conjunto de datos (una estructura sin el análisis suficiente),
- c) poseen el mismo nivel de conocimiento acerca del dominio del problema, contando con la misma documentación explicativa, y
- d) todos pueden ser considerados como expertos, sobre cómo realizar la transformación desde un conjunto de datos hasta llegar a una estructura de datos deseable;

entonces los arquitectos pueden realizar la transformación con cualquier método de razonamiento o con cualquier procedimiento lógico, pero todos siempre deben llegar al mismo resultado”.

Esto es lo que permitiría validar o no la hipótesis de trabajo.

3. RESULTADOS ESPERADOS

El proyecto formalmente inició en enero del corriente año, por lo cual no se dispone aún de resultados, pero ya se han enunciado líneas de trabajo y un cronograma a seguir para conseguir los avances previstos.

Es objetivo, del grupo de investigación, proponer un conjunto de buenas prácticas para el diseño del modelo conceptual de datos

normalizado, para bases de datos relacionales, tendientes a explicar las reglas y características que fundamentan que el modelo conceptual de una base de datos relacional puede ser único, para un dominio dado.

Los resultados esperados del proyecto se pretenden documentar y de esta manera lograr los siguientes objetivos:

- Determinar los principios que argumentan la hipótesis.
- Contrastar y ejemplificar estructuras validadas de modelos conceptuales con posibles estructuras no válidas.
- Ejemplificar prácticamente la hipótesis.
- Identificar las premisas teóricas y prácticas que sustentan la hipótesis.
- Identificar todas las estructuras típicas que se utilizan en universo de soluciones.
- Identificar las estructuras diferentes que ofrecen una solución equivalente, si las hay, mostrando las similitudes y discrepancias de esas estructuras.
- Explicar la forma en que implementan una estructura de datos válida los arquitectos de datos.
- Sugerir un método de trabajo que permita lograr una estructura deseable, en la búsqueda de estandarizar el proceso.
- Documentar la estandarización de las acciones identificadas, junto a las reglas.

4. FORMACIÓN DE RECURSOS HUMANOS

El equipo de trabajo está constituido por: a) un director de proyecto, b) un codirector, c) cinco docentes investigadores, dos en formación y tres de apoyo, d) dos estudiantes becarios y e) un graduado.

Todos los docentes, que integran el equipo de investigación, comparten su actividad académica en la cátedra de Gestión de Datos, de tercer nivel de la Carrera de Ingeniería en Sistemas de Información, donde es central la temática de Bases de Datos. En el contenido mínimo de la asignatura el diseño de una estructura de datos ocupa un espacio

preponderante y en el tiempo de desarrollo ocupa la primera mitad de un cuatrimestre.

El reciente graduado, Ing. en Sistemas de Información Agustín Carrasco, invitado a participar en el proyecto, ha sido ya integrante de anteriores Proyectos de Investigación y como estudiante logró ganar en dos ediciones consecutivas el Concurso de Programación que se realiza en la UTN-FRC.

Los estudiantes Bárbara Romina Urbano Moreno y Santiago Manuel Argüello están cursando actualmente el cuarto año de la Carrera de Ingeniería en Sistemas de Información e iniciando el camino en investigación como becarios.

5. BIBLIOGRAFÍA

[1] Reinoso, E.; Maldonado, C.; Muñoz, R.; Damiano, L.; Abrutsky, M. - Bases de Datos - Edit. AlfaOmega Editores - Argentina - 2012- ISBN: 978-987-1609-31-4.

[2] Gastañaga, I., Maldonado, C., Martínez Spessot, C. I., & Hintermeister, E. (2006). TecnoDB una Base de Datos Relacional y Prometeo un método de aprovechamiento de Metadatos y Generador de Consultas. In VIII WICC.

[3] Marciszack, M., Maldonado, C., Martínez Spessot, C. I., Muñoz, R., Navarro, A., Peretti, J. P., & Roggero, L. (2009). Prometeo: una herramienta para el aprovechamiento de metadatos de base de datos relacionales. In XI WICC.

[4] Vaca, P. A., & Abrutsky, M. (2009). Herramienta para consultas complejas orientada a usuarios finales. In XI WICC.

[5] Paz Menvielle, M. A., Cuevas, J. C., Damiano, L. E., Muñoz, R., & Quinteros, S. (2013, June). Generador automático de modelos de datos normalizados en bases de datos relacionales. In XV WICC.

[6] Muñoz Roberto Miguel, Maldonado Calixto, Damiano Luis Esteban, Romero María Soledad, Cuevas Juan Carlos, Quinteros Sergio Ramón, Guevara Andrea, Carrasco Agustín. MultiDB, Plataforma Web para acceder a

diferentes Sistemas de Gestión de Bases de Datos Relacionales. XVIII WICC (2016)

[7] Giarratano, Joseph - Riley, Gary - Sistemas Expertos Principios y Programación - Editorial Cengage Learning / Thomson Internacional, 2005 - ISBN 9789706860590

[8] Date, C.J. Introducción a los Sistemas de Bases de Datos. Volumen 1. Quinta Edición. USA, 1993. Edit. Addison Wesley Iberoamericana, Inc. ISBN: 0-201-51859-7.

[9] Elmasri, R. y Navathe, S. Fundamentos de Sistemas de Bases de Datos. 5ta Edición, 2007 Edit. Pearson. ISBN: 978-84-7829-085-7.