



Modelo seudomatemático para el diseño de bases de datos relacionales

Pseudo-Mathematical Model for Design of Relational Databases

José Edwin Aponte Blanco¹, Jhon Alexis Méndez Lara²

Para citar este artículo: Aponte, J. E.; Méndez, J. A. (2016). Modelo seudomatemático para el diseño de bases de datos relacionales. *TIA*, 4(2), pp.28-35.

ARTÍCULO DE INVESTIGACIÓN

Fecha de recepción:
27-11-2015

Fecha de aceptación:
16-05-2016

ISSN: 2344-8288

Vol. 4 No. 2

Julio - Diciembre 2016

Bogotá-Colombia

Resumen

El presente artículo busca mostrar los diferentes términos y nociones acerca del diseño de bases de datos, por medio de la metodología creada por el profesor de la Universidad Distrital, ingeniero John Jairo Londoño, y su aplicación en las teorías de bases de datos planteadas por Armstrong y Boyce-Codd. Para tal fin, se presenta cómo, por medio del análisis de un sistema a través de la Cadena Lógica del Negocio del Sistema (CLNS), se puede construir una base de datos que cumpla con los lineamientos de la teoría de normalización tradicional; el entendimiento de esta nueva e innovadora metodología permite, en gran medida, simplificar el proceso de diseño de bases de datos para el ingeniero experto, todo esto sin olvidar los fundamentos ya establecidos y comprobados para el mismo fin.

Palabras clave: bases de datos relacionales, dependencias funcionales, diseño de bases de datos, modelo relacional.

Abstract

This article shows the different terms and notions about database design using the methodology created by the Universidad Distrital professor, engineer John Jairo Londoño, and its application on the theories raised by Armstrong and Boyce-Codd. For this purpose, it is presented by the analysis of a system, through Logical Chain of Business System, how it can build a database which complies with guidelines of traditional normalization theory. The understanding of this new innovative methodology allows greatly simplify database design process to the expert engineer, all without forgetting the existing tested foundations for the same purpose.

Keywords: relational databases, functional dependencies, database design, relational model.

² Especialización en Ingeniería de Software, Universidad Distrital Francisco José de Caldas

INTRODUCCIÓN

El uso de modelos de entidad relación para dar explicación a las bases de datos, con referencia a un sistema de información, se ha convertido en un tema poco fácil de abordar al momento de definir si se realizan de la manera correcta o no, además de si cumplen con las especificaciones y requerimientos contemplados en el contexto del sistema; la falta de utilización de una metodología fuera de lo matemático y algebraico hace que los modelos se vuelvan un poco subjetivos al momento de analizarlos. El ingeniero John Jairo Londoño Pérez, profesor de la Universidad Distrital Francisco José de Caldas, es el creador de una metodología que resuelve lo anteriormente planteado, con el fin de obtener una vista general de la creación de un diseño de bases de datos a partir del contexto que da un sistema; esto se logra con la contemplación de dos preguntas: ¿para quién se quiere controlar? y ¿qué se quiere controlar? Respondiendo la primera, se tiene lo que el profesor denomina “el sujeto”, y la segunda, “el grupo transaccional”; ambos componentes forman lo que se denomina la Cadena Lógica de Negocio del Sistema (CLNS), que no es más que la abstracción completa del sistema a analizar.

Lo que se pretende en este documento es plantear los lineamientos generales de dicha metodología con el ánimo de orientar al lector, con una plataforma inicial, para construir un conocimiento más profundo en este tema.

BASES DE DATOS RELACIONALES

Las bases de datos relacionales son un tipo de base de datos que pretenden representar los datos en forma de tablas y relaciones, es uno de los modelos más utilizados y populares, fue creado por Edgar Frank Codd a finales de los años 60. Según el profesor Henry F. Korth en su libro *Fundamentos de bases de datos*, las bases de datos relacionales consisten en un conjunto

de tablas a las cuales se les asigna un nombre exclusivo, tienen como estructura la selección de atributos para cada tabla y un dominio asociado a ese atributo [2]. Estos modelos se basan en fundamentos matemáticos, ya que las relaciones de las tablas se ven como subconjuntos del producto cartesiano en la lista de dominios. Se utilizan los términos matemáticos relación y tupla en lugar de los términos tabla y fila; una variable tupla es una variable que representa a una tupla, en otras palabras, una tupla que representa al conjunto de todas las tuplas.

Modelo seudomatemático para el diseño de bases de datos relacionales

Es un modelo creado por el ingeniero John Jairo Londoño Pérez, muestra una metodología para que se diseñen las bases de datos a partir de dos únicas preguntas: ¿qué se quiere controlar? y ¿para quién se quiere controlar?, estas preguntas se tienen en cuenta para abstraer la forma más precisa el contexto del sistema al que se quiere diseñar una solución en bases de datos. Dicha metodología parte de los fundamentos matemáticos como son los axiomas de Armstrong, aplicables a las formas de normalización de las bases de datos, con la diferencia de que esta propone nuevos términos más accesibles que resumen, de alguna manera, el gran número de preguntas que se formulan originalmente dentro de los diferentes axiomas nombrados.

Dentro de los términos soportados por la metodología se tiene al sujeto, el cual responde a la pregunta inicial: ¿para quién se quiere controlar?, y el grupo transaccional que responde a: ¿qué se quiere controlar?

Con base en lo anterior, se propone realizar una cadena lógica del negocio que permite, a partir de las dependencias funcionales básicas, alinear todo lo concerniente a la ingeniería del software y renueva, de alguna manera, el pensamiento hasta hoy utilizado para contextualizar la problemática de los sistemas de información [1].

Se proponen entonces cinco esquemas para diseñar las bases de datos que se utilizan de acuerdo al contexto que de un sistema, todas ellas basadas en la relación que existe entre un sujeto y su grupo transaccional; para poder identificar dichas relaciones se deben tener en cuenta las siguientes preguntas:

- ¿Existe dependencia funcional, exclusiva del grupo transaccional, respecto al sujeto?
- ¿Existe dependencia funcional no exclusiva del grupo transaccional respecto al sujeto?
- ¿Qué variables dependen funcional y estrictamente de los dos componentes de la llave?
- ¿Qué variables dependen funcional y estrictamente del segundo componente de la llave?

Cadena Lógica del Negocio del Sistema (CLNS)

La CLNS permite determinar la dimensión del producto de software que se pretende construir y, en consecuencia, deberá alinear con todo componente que se involucre en su desarrollo [1]; en lo que atañe al diseño de la base de datos, es a partir de la cadena del negocio como se hace posible aplicar los conceptos de las dependencias funcionales, resumidos en las cuatro preguntas de la metodología del modelo seudomatemático hablado. La CLNS se abstrae del contexto del sistema, no permite variaciones en su estructura ya que lo que reconoce precisamente es el modelamiento del comportamiento del sistema en una realidad a partir de la definición de sujetos y grupos transaccionales.

Sujeto

Teniendo en cuenta la teoría del modelo seudomatemático para el diseño de las bases de datos, se hace referencia a la definición de un sujeto para poder diseñar una base de datos

de acuerdo al contexto de un sistema; el sujeto se presenta como una entidad que responde a la pregunta: ¿para quién se quiere controlar? [1]. Respondiendo a dicha pregunta, el sujeto definido se presenta como principal entidad de la cadena lógica del negocio que se define del contexto del sistema, que no es más que una abstracción de una realidad; cuando es evidente en un diseño de bases de datos para qué entidad se va a controlar, se puede definir con claridad qué se quiere controlar exactamente por medio de ese sujeto.

Grupo transaccional

Los grupos transaccionales, como respuesta a la pregunta: ¿qué se va a controlar?, determinan otra entidad que conforma la cadena lógica del negocio. Se habla de grupos ya que pueden ser muchos los que son controlados por un solo sujeto en la CLNS [1]; en consecuencia para una CLNS se puede tener un sujeto con un grupo transaccional, un sujeto con dos grupos transaccionales independientes, un sujeto con un grupo transaccional anidado, un sujeto con dos grupos transaccionales independientes y subgrupos transaccionales para cada grupo inicial y un sujeto que contenga dos grupos transaccionales independientes y uno de ellos contiene un subgrupo transaccional. Todas estas combinaciones se pueden dar en una CLNS, la selección de cada una dependerá del análisis del contexto del sistema.

Dependencia funcional

En un diseño de bases de datos existen relaciones que sirven para conectar una o más entidades; dichas conexiones se establecen entre los atributos de cada entidad. Para el modelo seudomatemático las relaciones se dan por medio de los atributos del sujeto y el grupo o los grupos transaccionales; las dependencias funcionales juegan un papel importante en la definición de las relaciones entre entidades, ya que estas establecen el

grado de dependencia que tiene un sujeto con un grupo transaccional; se hace posible aplicar los conceptos de las dependencias funcionales en el modelo seudomatemático resumido en cuatro preguntas principales [1], las cuales son:

- ¿Define la llave primaria del grupo transaccional una dependencia funcional exclusiva de la llave primaria del sujeto?
- ¿Define la llave primaria del grupo transaccional una dependencia funcional no exclusiva de la llave primaria del sujeto?
- ¿Qué variables del grupo transaccional definen una dependencia funcional estricta de la llave primaria del sujeto, más la llave primaria del grupo transaccional?
- ¿Qué variables del grupo transaccional definen una dependencia funcional estricta de la llave primaria del grupo transaccional?

Dependencia funcional exclusiva

Cuando se tienen relaciones entre entidades que requieran que un grupo transaccional dependa directamente de un sujeto, se presentan dependencias funcionales exclusivas; este término es propio del modelo seudomatemático creado por el ingeniero Londoño para referenciar los atributos del ente transaccional con respecto al sujeto, está dada, en comparación con la teoría de bases de datos, como una relación de uno a muchos.

Dependencia funcional no exclusiva

Cuando se establece en la CLNS que un grupo transaccional depende del sujeto directamente y ese sujeto también tiene una dependencia con el grupo transaccional, se determina que estas dos entidades tienen una dependencia funcional no exclusiva; en comparación con la teoría de bases de datos estas relaciones se ven como relaciones de muchos a muchos, con el modelo seudomatemático se resuelven siempre con la llave compuesta [1].

APLICACIÓN DEL MODELO SEUDOMATEMÁTICO EN LA TEORÍA DE NORMALIZACIÓN DE LAS BASES DE DATOS

Para efectos de ilustración, se lleva a cabo la solución de un problema sencillo que consiste en diseñar el modelo de bases de datos que permita el almacenamiento de la información de clientes, facturas y productos. El proceso de normalización se define a través del cumplimiento de varias etapas conocidas como formas normales.

Primera forma normal

Una tabla se encuentra en primera forma normal cuando:

- a) Los atributos de la tabla son atómicos, es decir, no tienen posibilidad de ser más divididos; por ejemplo, el nombre de una persona solo se puede separar en nombres y apellidos, Tabla 1.
- b) La tabla tiene definido un campo que permite identificar de forma única e inequívoca todo el registro; este campo se conoce como clave principal o llave primaria (*primary key*), Tabla 2.

Se puede apreciar, en las figuras 1 y 2, que cada tabla ahora cuenta con sus atributos separados y la asignación de una llave primaria candidata.

Segunda forma normal

Los atributos de la tabla que no son llave principal tienen correspondencia única con la llave principal, por ejemplo, el nombre de una persona tiene correspondencia única con su cédula (Figura 3).

Figura 1. Tabla sin formas normales

Cedula	Nombre	Ciudad	Factura	Fecha	Producto	Cantidad	Vr Unitario
123	Juan Pérez	Bogotá	1	15/07/15	Teclado	2	15000
123	Juan Pérez	Bogotá	1	15/07/15	Mouse	1	14000
234	Daniel Mendoza	Bogotá	2	16/07/15	Monitor	1	285000
234	Daniel Mendoza	Bogotá	2	16/07/15	Teclado	1	15000
234	Daniel Mendoza	Bogotá	2	16/07/15	Mouse	1	14000
345	Maria Perdomo	Cali	3	17/07/15	Monitor	1	285000
456	Claudia Forero	Villavicencio	4	18/07/15	Monitor	1	285000
456	Claudia Forero	Villavicencio	4	18/07/15	Teclado	1	15000
456	Claudia Forero	Villavicencio	4	18/07/15	Mouse	1	14000
456	Claudia Forero	Villavicencio	4	18/07/15	Parlantes	1	12000
456	Claudia Forero	Villavicencio	4	18/07/15	CPU	1	850000

Fuente: elaboración propia.

Figura 2. Tablas en primera forma normal

Cedula	Nombre	Apellido	Ciudad
123	Juan	Pérez	Bogotá
234	Daniel	Mendoza	Bogotá
345	Maria	Perdomo	Cali
456	Claudia	Forero	Villavicencio

Ciudad
Bogotá
Cali
Villavicencio

Factura	Fecha
1	15/07/15
2	16/07/15
3	17/07/15
4	18/07/15

Producto	Vr Unitario
Teclado	15000
Mouse	14000
Monitor	285000
Parlantes	12000
CPU	850000

Fuente: elaboración propia.

Figura 3. Tablas en segunda forma normal

Cedula	Nombre	Apellido	CodCiudad
123	Juan	Pérez	1
234	Daniel	Mendoza	1
345	Maria	Perdomo	2
456	Claudia	Forero	3

CodCiudad	Ciudad
1	Bogotá
2	Cali
3	Villavicencio

CodProducto	Producto	Vr Unitario
1	Teclado	15000
2	Mouse	14000
3	Monitor	285000
4	Parlantes	12000
5	CPU	850000

Factura	Fecha
1	15/07/15
2	16/07/15
3	17/07/15
4	18/07/15

Fuente: elaboración propia

Tercera forma normal

Los atributos no clave deben ser completamente dependientes de la clave principal; para el cumplimiento de esta forma normal se crean las tablas que relacionan clientes con facturas y facturas con productos (Figura 4).

Cuarta forma normal

Los atributos que pueden tener multivalores se separan en dos o más relaciones independientes; el ejemplo se aplica para el caso de muchos productos en una factura y un producto en varias facturas (Figura 5).

De cara al modelo del ingeniero Londoño se identifican los elementos de la CLNS como sigue (Figura 6).

Figura 4. Tablas en tercera forma normal

Cedula	Nombre	Apellido	CodCiudad
123	Juan	Pérez	1
234	Daniel	Mendoza	1
345	Maria	Perdomo	2
456	Claudia	Forero	3

CodProducto	Producto	Vr Unitario
1	Teclado	15000
2	Mouse	14000
3	Monitor	285000
4	Parlantes	12000
5	CPU	850000

CodCiudad	Ciudad
1	Bogotá
2	Cali
3	Villavicencio

Factura	Fecha	Cedula	Productos
1	15/07/15	123	1,2
2	16/07/15	234	3,1,2
3	17/07/15	345	3
4	18/07/15	456	3,1,2,4,5

Fuente: elaboración propia

Figura 5. Tablas en cuarta forma normal

Cedula	Nombre	Apellido	CodCiudad
123	Juan	Pérez	1
234	Daniel	Mendoza	1
345	Maria	Perdomo	2
456	Claudia	Forero	3

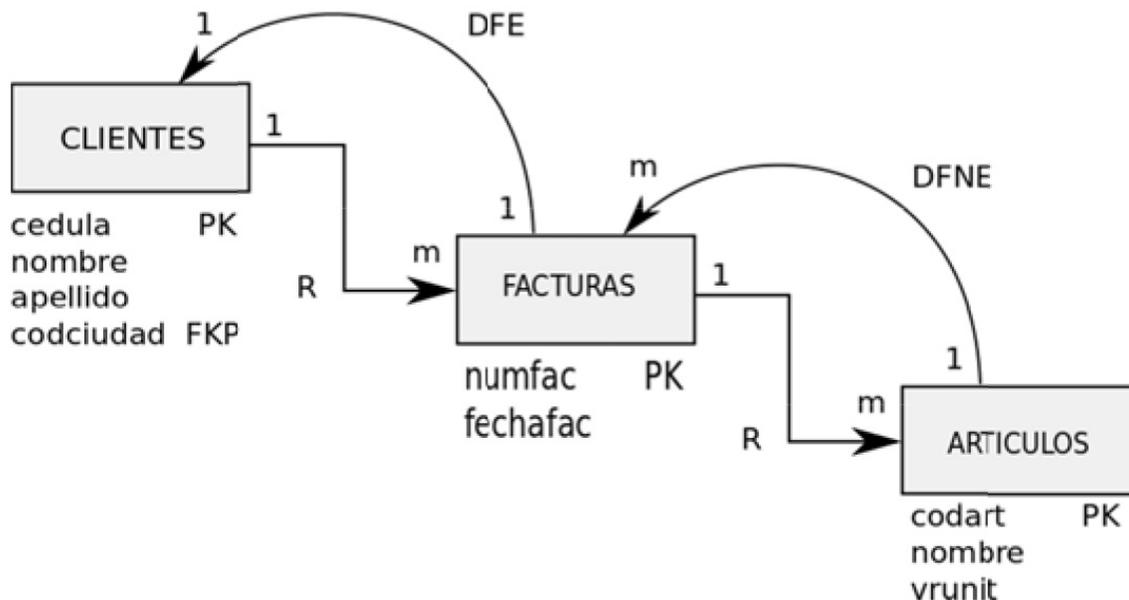
CodProducto	Producto	Vr Unitario
1	Teclado	15000
2	Mouse	14000
3	Monitor	285000
4	Parlantes	12000
5	CPU	850000

CodCiudad	Ciudad
1	Bogotá
2	Cali
3	Villavicencio

Factura	Fecha	Cedula
1	15/07/15	123
2	16/07/15	234
3	17/07/15	345
4	18/07/15	456

Factura	CodProducto	Cantidad	Vr Unitario
1	1	2	15000
1	2	1	14000
2	3	1	285000
2	1	1	15000
2	2	1	14000
3	3	1	285000
4	3	1	285000
4	1	1	15000
4	2	1	14000
4	4	1	12000
4	5	1	850000

Fuente: elaboración propia

Figura 6. Cadena lógica del negocio

Fuente: elaboración propia

Desde el análisis que se puede hacer para confrontar el cumplimiento del modelo propuesto por el ingeniero Londoño de las formas normales, se encuentra que: el cumplimiento de la primera y la segunda forma normal se visualizan a través de las siguientes condiciones:

- En la definición de los atributos de cada sujeto o grupo transaccional, se definen como atómicos.
- En cada uno se define un atributo como clave primaria PK.

El cumplimiento de la tercera y cuarta forma normal se materializa en cuanto se cumple que:

- Los atributos son completamente dependientes de la PK.
- Los atributos definidos como Foreign Key Default (FKD) permiten relacionar las tablas para dependencias funcionales exclusivas y no exclusivas.

CONCLUSIONES

Este modelo planteado por el ingeniero Londoño es nuevo e innovador, permite al ingeniero experto en diseño de base de datos tener una vista general del sistema para coordinar el diseño de software con el de la base de datos: propone simplificar el modelamiento de bases de datos, pues al iniciar los pasos de esta se aplican directamente los principios de normalización. La metodología no renuncia a los fundamentos tradicionales de las bases de datos, sino que concibe la teoría de Armstrong como un referente en la definición de la lógica del sistema.

Entendiendo que la metodología ofrece simplicidad, inmediatez, eficacia y confiabilidad, se hace necesario pensar en el esbozo de una herramienta que permita apoyar el diseño de bases de datos contemplando la teoría anteriormente descrita; la idea de construir el cuerpo práctico del

modelo seudomatemático, para el diseño de bases de datos relacionales propuesto por el ingeniero John Jairo Londoño Pérez, es un buen comienzo en la búsqueda de construcción de conocimiento alrededor de los nuevos desarrollos que se vislumbran a partir de esta metodología.

REFERENCIAS

- [1] Londoño J. (2011). *Modelo seudomatemático para el diseño de las bases de datos relacionales*.
- [2] Silberschatz, A., Korth H. y Sudarshan S. (2002). *Fundamentos de bases de datos*. Madrid: McGraw-Hil.