

EL CONCEPTO DE BASES DE DATOS RELACIONAL UNA MIRADA DESDE LO EPISTEMOLÓGICO Y LO MATEMÁTICO

María Dora Alba Sánchez Gómez, Juan Evangelista Gómez Rendón

Escuela de ciencias básicas, tecnología e ingeniería, Universidad Nacional Abierta y a Distancia

CEAD Medellín, Colombia

mariadora.sanchez@unad.edu.co

fuanevangelista@yahoo.com

Resumen— El artículo muestra una simbiosis armónica entre el diseño, la teoría de conjuntos, las relaciones, las funciones, la lógica matemática y cómo a pesar de que las “bases de datos relacionales” se fundamentan en unas teorías matemáticas sólidas y sencillas, muchas de las conceptualizaciones expresadas en los textos y en el ciberespacio, se abordan desde una perspectiva meramente técnica; solo se interesan por el manejo rápido de la herramienta, dejando de lado su análisis y diseño conceptual, de tal manera, que toda la versatilidad, la robustez y la potencialidad con que cuenta en este nuevo milenio este poderoso software se pueda perder o simplemente subperfiar por un mal análisis o por un defectuoso diseño.

Palabras claves— Bases de datos relacionales, teoría de Conjuntos, diseño conceptual, relaciones, funciones, lógica matemática, software de bases de datos.

Abstract— The article shows a harmonic symbiosis between design, set theory, relations, functions, mathematical logic, and how even though the "relational database" is based on a robust and simple mathematical theories, many of the conceptualizations expressed in texts and in cyberspace, are approached from a purely technical perspective, only interested in the quick handling of the tool, leaving out his analysis and conceptual design, so that all the versatility, robustness and the potential it has in this new millennium, this powerful software can be lost or just a bad subprofiles by a faulty analysis or design.

Keywords— Relational databases, set theory, conceptual design, relations, functions, mathematical logic, database software.

INTRODUCCIÓN

El ánimo que estimula la creación de este artículo, es mostrar cómo en algunos textos consultados en libros o en el ciberespacio, se parte de unas definiciones erradas de lo que es una base de datos relacional teniendo en cuenta que una buena conceptualización se inicia con definiciones claras, correctas, sin ambigüedades y ojalá amigables, y en muchos de estos casos cambian la verdadera y profunda dimensión que tiene

esta tecnología para la estabilidad y consistencia de los sistemas de información en las organizaciones y que además les permita y asegure permanencia en el espacio tiempo aunque los datos aumenten significativamente.

Por otro lado, durante los cursos de “bases de datos básicos” impartidos a través de muchísimos años en varias universidades e instituciones del país, comenzamos a observar cómo muchos de los estudiantes utilizaban indistintamente las diversas concepciones o terminologías de “las bases de datos”, confundiendo, hasta tal punto que le daban más importancia y trascendencia al manejo de la herramienta que al diseño conceptual de éstas. Pero, una vez que los viajeros del curso entraban a profundizar y a socializar el análisis y el diseño de éstas, se iban dando cuenta de la relevancia que tienen el diseño y el análisis sobre la utilización y el debido control de la herramienta. Comenzaban a tener más criterio sobre la elección de un texto u otro, de tal manera que se volvían unos críticos constructivos bien argumentados e inclusive no importando la terminología utilizada podían diferenciar una de otras para no confundirse.

Es fundamental entonces socializar que para implementar una “base de datos relacional” se deban:

- analizar los datos de la empresa, sus flujos, necesidades, puntos críticos, falencias y requerimientos de la información.
- Realizar un análisis concienzudo que permita elegir el diseño más flexible, de tal forma, que garantice a través del espacio tiempo la consistencia de la información aunque los datos de la empresa aumenten
- tener un criterio válido para la elección del software y la plataforma tecnológica para proseguir con la implementación y la administración objetiva para ponerla operativa y a disposición de la organización.

CONCEPTUALIZACIÓN DE LAS BASES DE DATOS

Hoy en día, comienzo de la segunda década del siglo XXI, el llamado siglo de la información, se habla de las bondades, potencialidades, tendencias, recursos, poder, necesidad,

M. D. Sánchez labora en la Universidad Nacional Abierta y a Distancia desde 2009-1, (e-mail: mariadora.sanchez@unad.edu.co).

J.E. Gómez Rendón labora en la Universidad Nacional Abierta y a Distancia desde 2005-2 (e-mail: fuanevangelista@yahoo.com).

transportabilidad, disponibilidad, de las bases de datos, o mejor dicho, de las herramientas de las bases de datos. De hecho, muchos definen las bases de datos en término de las herramientas tecnológicas que se van desarrollando y que se van ofertando en los mercados; y es, justo aquí, donde comienzan a aparecer los problemas o, los reparos para conceptualizarlas. Más aún, hacen abstracción de la teoría matemática que hay detrás de este concepto y por consiguiente al análisis y al diseño conceptual de éstas no se les da la importancia ni la relevancia que merecen; solo se reduce a una simple creación o aun diseño de estructuras planas de datos específicos en un software particular; es decir, se propone un viaje casi directo (sin el encanto y sin la fortaleza de los esenciales conceptos) hacia el manejo y administración de una herramienta específica así ella sea muy poderosa.

Por otra parte, se observa cada vez más que las personas se han vuelto amantes y adictas a las consultas por la gran red, pocas analizan y profundizan en los contenidos debido en parte a la vertiginosidad con que necesitamos o se procesa la información o porque estamos viviendo unos momentos de rapidez para todo; es un mundo donde las imágenes pueden más que las palabras y éstas nunca argumentan, solo muestran y tratan de sorprender o de impactar y es así cómo el problema se acentúa más, pues si bien hay libros que son muy buenos referentes, también pueden presentar problemas a la hora de definir o socializar los términos, aunque en menor grado. Para sustentar lo anterior, miremos algunas definiciones o concepciones que se encuentran en algunos textos o viajando y consultando indistintamente en el ciberespacio.

- 1) las bases de datos son una serie de archivos relacionados". (1)
- 2) "las bases de datos son una fuente central de datos que está pensada para ser compartida por muchos usuarios". (2)
- 3) "las bases de datos son un sistema computarizado para guardar registros". (3)
- 4) "las bases de datos son un conjunto de datos persistentes". (4)
- 5) "un gestor de bases de datos, es una colección de datos interrelacionados y un conjunto de programas para acceder a dichos datos". (5)

Si observamos las definiciones 2, 3 y 5; están enfocadas a las herramientas y por consiguiente la persona que desea saber sobre el tema de las bases de datos relacionales, cree que solo es suficiente con aprender una herramienta y peor aún, alejadas de toda la concepción teórica que las originó.

Ahora, analicemos las otras dos definiciones, las dos están a medias; la número uno, si bien introduce el concepto de

interrelación, la define en términos de unas estructuras de datos (archivos de datos), que se utilizan mucho en la "tercera generación"⁶ y recordemos que el concepto "bases de datos relacionales", es un producto de la "cuarta generación"⁷ y por consiguiente se tiene el problema de que este concepto es una simple extensión de archivos, lo cual es un error.

La cuarta definición, habla de un conjunto de datos, mas no establece el término interrelacionados, que es una de las grandes bondades y precisiones que encierra este concepto.

Ahora miremos analizar unas definiciones que se encuentran en INTERNET, donde se puede apreciar fácilmente el desconocimiento epistemológico y matemático de este importante concepto:

"las bases de datos es una colección de información, que son guardadas en un medio magnético"

"las base de datos son un conjunto de datos o información relacionada, los cuales son almacenados en un computador"

"las bases de datos son una serie de información almacenadas para su posterior consulta"

En las anteriores definiciones lo primero que tenemos que entrar a socializar es el concepto diferenciador entre "datos" e "información". El término "**dato**" viene del latín DATUM que significa hecho, algo que podemos percibir o sentir con nuestros sentidos (oído, vista, olfato, tacto y gusto), pero que no necesariamente nos signifique algo. Por otra parte, el concepto "**información**" es más significativo; de tal forma que muchos autores lo definen como "un dato con significado" y he aquí la gran diferencia y complejidad de este término, pues lo que hoy le puede ser significativo a una persona, mañana quizá no lo sea; es decir se torna simplemente en un dato. Adicional a esto el concepto de significación le atañe única y exclusivamente al ser humano, es él, en última instancia quien puede interpretar qué datos son información para él; es más ni siquiera otra persona puede decir con absoluta certeza que esto es información para otra persona. Sin embargo, es bueno aclarar que los datos son la materia prima y formal de la información y que por lo tanto sin datos no hay información, pero no implica que todo dato sea información. Fundamentados en el análisis anterior, podemos entonces afirmar que "las bases de datos, son estructuras de datos y no de información; ellas solo almacenan hechos".

Por otra parte, debemos comprender y socializar el significado de los términos "**conjunto**", "colección", "serie".

Desde la definición de la Real Academia Española de la Lengua; "**conjunto**" es la totalidad de los elementos o cosas poseedores de una propiedad común, que los distingue de

¹ EFFY-OZ, Administración de sistemas de información. Pag.30

² KENDALL & KENDALL. Análisis y diseño de sistemas. Pag.40

³ C.J. DATE. Introducción a los sistemas de bases de datos. Pag. 5

⁴ Ibidem.

⁵ SILBERSCHATZ, KNORT,SUDARSHAN. Fundamentos de bases de datos. Pag. 1.

⁶De 1965 a 1975 se utilizaron estos lenguajes donde cada vez que se deseaba utilizar datos almacenados en un programa se debía definir la estructura de los archivos. Es el caos de los lenguajes COBOL, FORTRAN.

⁷A partir del 80 se comienzan a utilizar estos lenguajes no procedimentales en los cuales se usaban datos centralizados a través de las herramientas de las bases de datos. Es el caso de ORACLE y DB2.

otros; “colección” es el conjunto ordenado de cosas, por lo común, de una misma clase y reunidas por su especial interés o valor. Como podemos observar estos términos son parecidos, pero se difieren entre sí en que el conjunto los elementos no deben estar ordenados y esto hace una gran diferencia. Además, el término “conjunto” visto desde las matemáticas, se define como los entes matemáticos que tienen una propiedad en común, volvemos y observamos que los elementos no tienen que estar ordenados.

Ahora analizando el término “serie” nos encontramos que es expresión de la suma de los infinitos términos de una sucesión, donde definitivamente difiere de los dos primeros (conjunto y colección). Realizando un análisis concienzudo de lo anterior, desde la perspectiva matemática, las bases de datos son un conjunto de datos, que no necesariamente deben estar ordenados (como no lo son los elementos de un conjunto cualquiera), por lo que el término “colección” queda abolido y con más razón el término “serie”.

Siendo así, entonces podemos tener una definición más acorde, más significativa y más cercana con el sustento teórico matemático de las bases de datos:

Es un conjunto de datos relacionados .Cabe recordar que cuando un conjunto se determina por comprensión es porque sus elementos se definen de acuerdo a una característica común a todos ellos; por ejemplo el conjunto de las vocales.

La pregunta ahora es, ¿por qué relacionales?, bueno ya se advertía anteriormente que unas de las grandes bondades de las bases de datos (cuarta generación “4GL”) con respecto a los archivos de datos (tercera generación “3GL”), era la de poder relacionar los datos unos con otros. Ahora bien, cabe resaltar que todas las operaciones básicas de acceso a la base de datos (álgebra relacional), tiene su sustento en las operaciones básicas de conjuntos (unión, intersección, diferencia, complemento), que es lo que en última instancia motivó a los creadores de las “bases de datos relacionales” a apoyarse en las matemáticas; mas propiamente, en la teoría de conjuntos, y ésto es lo que ha permitido que de las bases de datos se hable mucho, porque la creación de las herramientas y su utilización, están afianzadas en una muy buena teoría matemática siendo así más comprensibles y amigables para el ser humano lo cual posibilita su entendimiento, manejo y socialización en muchos campos del conocimiento

.Las bases de datos relacionales también se apoyan en el concepto matemático de “función”, el cual en mundos de dos dimensiones no permiten que existan dos pares o parejas ordenadas con la primera componente igual. Ello impediría por ejemplo que en una base de datos que se interese por los números de cédula y nombres y apellidos de una persona, por ejemplo, se diese la posibilidad de que para un mismo número de cédula se tuviesen dos nombres diferentes. Incluso socialmente la validez de esta propuesta no es reconocida.

SOLUCIÓN DE CASO “EMPLEADOS”

Veamos un ejemplo de la vida real, tomemos el caso de “empleados” en el cual se quiere crear una base de datos de personal para una compañía, con las siguientes características:

La compañía tiene un “conjunto” de departamentos. Cada departamento tiene un conjunto de empleados, un conjunto de proyectos y un conjunto de oficinas. Los proyectos tienen empleados asignados, y las oficinas empleados que laboran en ellas. Los cargos están codificados, de tal forma, que tienen un código único y una descripción del cargo. Cada empleado tiene una historia de cargos ocupados, en la cual se requiere saber la fecha en que ha ocupado éstos.

La base de datos debe contener o manejar la siguiente información: Por cada departamento se necesita, código de departamento (único), nombre, presupuesto anual. Por cada empleado se necesita, código de empleado (único), nombre, dirección y teléfono de residencia, departamento y proyecto al que está asignado actualmente, número de oficina y cargo actual. Por cada proyecto se requiere número de proyecto (único), nombre, presupuesto del proyecto y departamento al que pertenece. Por cada oficina, código de oficina (único), área en metros cuadrados, números de teléfono.

Además se sabe que, un empleado de la empresa pertenece a lo máximo un departamento al tiempo. Un empleado está vinculado a lo máximo a un proyecto al tiempo. Un empleado tiene a lo máximo una oficina al tiempo. Un empleado desempeña a lo máximo un cargo al tiempo. Una oficina pertenece a lo máximo un departamento al tiempo.

Para dar una buena solución a este caso, primero debemos identificarlas “entidades”. Observando y analizando objetivamente la información del caso, encontramos los siguientes sustantivos en plural: Departamentos, Empleados, Proyectos, Oficinas, Cargos e Historia de los Cargos.

ENTID AD	Dpt os.	Em pl.	Pro y.	Ofic.	Carg os	Hist_Ca rg.
Dptos.	x	1-N	1-N	1-N	X	X
Empl.	1-1	X	1-1	1-1	1-1	1-N
Proy.	1-1	1-N	X	X	X	X
Ofic.	1-1	1-N	X	x	X	X
Cargos	X	1-N	X	X	x	1-N
Hist_Ca rg.	X	1-1	X	X	1-1	x

Figura 1 – Matriz relación caso “empleados”

En la matriz anterior, observemos las relaciones y cardinalidades de Historias de Cargos. Lo primero es que solo está relacionada con las entidades Empleados y Cargos. Miramos que un empleado puede tener como máximo varias historias de cargos, y esto porque una persona durante la vida laboral puede desempeñar, obviamente, varios cargos; y una historia de cargos le pertenece máximo a un empleado y debido a que cada vez que un empleado cambia de cargo, se genera una instancia u ocurrencia en su historia de cargos. Por otro lado, con respecto a la relación que tiene con “cargos”, una “Historia de Cargo”, tiene máximo un Cargo, y un Cargo, puede encontrarse en varias instancias u ocurrencias de “Historia de Cargos”, debido a que pueden existir varios “empleados” con el mismo cargo. Con respecto a las demás “relaciones”, son claros y sociales sus lineamientos, de acuerdo al enunciado del caso especificado.

EXPLICACIÓN MATEMÁTICA DEL CASO ANTERIOR

Las bases de datos relacionales son una filosofía, que nos estimula y obliga a los profesionales encargados del diseño de ellas, a mirar más hacia la realidad del mundo empresarial, del mundo científico, del mundo del conocimiento, del mundo de la tecnología, si queremos propender que este recurso informático sea el sustento para almacenar, soportar y generar información en un medio donde ésta es la base del poder.

En las bases de datos relacionales y tal como lo ilustra el caso propuesto en la matriz que soporta el trabajo de la información, se percibe claramente que el lenguaje formal nos conduce inexorablemente al concepto de **relación**.

Las relaciones presentes en la matriz del inicio de la solución del caso permiten visualizar y asociar un conjunto de datos, muchos de los cuales están visiblemente interrelacionados entre sí generando de forma espontánea el concepto de **tupla** el cual supera en contenido y en dimensiones el concepto formal del plano cartesiano de **pareja ordenada** en el cual las **tuplas** son la materia prima de la Geometría Analítica. Las tuplas entonces dimensionan el pensamiento humano y lo ayudan a superar las limitaciones inevitables de los planos cartesianos normales donde el concepto de dimensión definido por el álgebra lineal es de dos: es decir en ese mundo solo se tienen conjuntos de dos vectores linealmente independientes y los cuales son capaces de generar cualquier vector de ese espacio vectorial.

Por otro lado, es bueno recordar que toda **función** es una relación, pero no toda relación es una **función**: por eso en la matriz relación anterior observamos que la **cardinalidad** nos ayuda a identificar si estamos ante una función o no, así: tomemos la relación que va de la entidad *Departamentos* a *Empleados* (encima de la diagonal), ésta es de 1-N, lo cual significa que un departamento (1) puede tener varios empleados (N). Ahora si tomamos la relación en el sentido contrario, es decir, de *Empleados* a *Departamentos* (debajo de la diagonal), observamos que la cardinalidad es de 1-1, lo cual

significa que un empleado (1) solo pertenece a un departamento (1). De lo anterior podemos concluir, que toda relación es en doble sentido y genera como resultado, que la relación entre *Departamentos* y *Empleados* es de 1,1 – 1,N, simplificando, de 1-N, lo cual es una función.

Invitamos ahora a los lectores de este artículo a que se dediquen a volar por espacio tiempos de muchas dimensiones y al hacerlo irán descubriendo que el llegar a otro mundo simplemente encuentra conjuntos de datos que le van mostrando la sencillez o complejidad de sus estructuras.

Para los neófitos en la comprensión y análisis práctico o empírico, de las bases de datos relacionales los pedagogos o los amigos de la propedéutica muestran que el aprendizaje de la herramienta se facilita jugando con una estrategia mental muy asimilable que es la interpretación de las relaciones existentes como si ella fuese una tabla matricial compuesta por registros (tuplas) y columnas (campos). Sin embargo desde el principio planteamos que el ánimo del escrito era ir más allá del aspecto meramente herramienta y mostrar que tras una base de datos relacional hay un tejido maravilloso, amigable, social, de conceptos matemáticos formales que permiten soportar toda una teoría fuerte y estable cuya solidez y logros han aportado inmensas cantidades de información a los seres humanos y han hecho posible que el mundo sea cada vez más pequeño y comprensible para todos.

Como conclusión se puede decir, que el concepto de “bases de datos relacionales”, además de ser una poderosa herramienta tecnológica, tiene una sólida y formidable aplicación matemática soportada amigablemente en los conceptos básicos de conjuntos, relaciones, funciones y los cuales a su vez son componentes importantes de un curso formal de “lógica matemática”.

Esperamos que las bases de datos sigan mostrando o generando información valiosa y apropiada que nos ayude a superar las enormes y profundas brechas que nos separan en un mundo de mercados fugaces y voraces y que no han permitido aún que el planeta tierra y más adelante el universo, sean escenarios democráticos y equitativos.

REFERENCIAS

Bibliografía

- Silberschatz y otros, *Fundamentos de bases de datos*, vol. I. New York: Wiley, 1950, p. 1.
- C.J. Date, “Introducción a los sistemas de bases de datos”, vol I. Madrid: Addison Wesley, 1993, p. 5
- Effy Oz, “Administración de sistemas de información”, Madrid: Prentice Hall, 2001, p. 30.
- Kendal y Kendal, “Análisis y diseño de sistemas”, Mexico: Prentice Hall, 2006, p. 40

BIOGRAFÍA



María Dora Alba Sánchez Gómez Se graduó en Ingeniera de Sistemas de la universidad EAFIT y es magistra en administración de empresa de la universidad del Norte. Su experiencia profesional incluye siete años como analista programadora en los ingenios azucareros del Valle del Cauca; seis años como directora de proyectos informático en el sector cooperativo financiero; seis años como directora del programa de tecnología e ingeniería de sistemas y más de veinte años de experiencia como catedrática en cursos de bases de datos, sistemas de información, gerencia de proyectos, planeación de sistemas de información. Actualmente es docente auxiliar de la Universidad Nacional Abierta y a Distancia.



Juan Evangelista Gómez Rendón Se graduó como “Físico Puro” de la Universidad de Antioquia y es “Especialista en ciencias electrónicas e informática” de la misma universidad. Creó el “paquete bananero” para las fincas de Jorge Ochoa en la región de Urabá y ha sido asesor externo de varias universidades para la implementación de las “tecnologías en informática y en electrónica”. Es el fundador de los “semilleros de investigación” del área de sistemas en el proceso de articulación de la media técnica con el Tecnológico de Antioquia. Ha trabajado en varias instituciones en los campos de algoritmos, sistemas de información y bases de datos básicos, entre otros. Actualmente se desempeña como tutor de la UNAD.